2/

REPORT RESUMES

ED 002 934 24
DEVELOPMENT OF AUDITORY PERCEPTION OF MUSICAL SOUNDS BY CHILDREN IN THE FIRST SIX GRADES.
BY- PETZOLD, ROBERT G.
WISCONSIN UNIV., MADISON
REPORT NUMBER CRP-766 PUB DATE 30 JUN 60
CONTRACT OEC-SAE-8411
EDRS PRICE MF-$0.18 HC-$5.12 128P.

DESCRIPTORS- *AUDITORY PERCEPTION, *MUSIC ACTIVITIES, *ELEMENTARY SCHOOL STUDENTS, *MUSIC EDUCATION, READING READINESS, GROWTH PATTERNS, RATE OF LEARNING SCORE, MADISON, WISCONSIN

THE AUDITORY PERCEPTION OF MUSICAL SOUNDS BY A SAMPLE OF 600 CHILDREN IN THE FIRST 6 GRADES WAS STUDIED. THREE TESTS WERE CONSTRUCTED FOR THIS STUDY. THEIR CONTENT WAS BASED UPON AN EXTENSIVE ANALYSIS OF TONAL AND RHYTHMIC CONFIGURATIONS FOUND IN THE SONGS CHILDREN SING. THE 45-ITEM AND ONE OF THE 20-ITEM TESTS WERE DESIGNED TO COLLECT DATA REGARDING THE AUDITORY PERCEPTION OF COMMON TONAL CONFIGURATIONS. FOR EACH OF THESE TESTS, THE SUBJECT HEARD AN AURAL PRESENTATION OF THE TONAL CONFIGURATION AND THEN WAS ASKED TO SING AN IMMEDIATE RESPONSE TO THE STIMULUS. THE ACCURACY OF THIS RESPONSE WAS THEN INTERPRETED AS A MEASURE NOT ONLY OF VOCAL CONTROL, BUT OF THE LEVEL AT WHICH AUDITORY PERCEPTION FUNCTIONED. IT WAS FOUND THAT AURAL UNDERSTANDING, WHICH IS A REFLECTION OF ACCURATE AUDITORY PERCEPTION, RESULTS FROM THINKING AND NOT FROM MECHANICAL IMITATION. JUDGMENTS WERE MADE INDEPENDENTLY BY THE CHILD AND BASED UPON HIS UNDERSTANDING OF BASIC MUSICAL CONCEPTS AND NOT JUDGMENTS MADE FOR THE CHILD BY SOMEONE ELSE. MUSIC EDUCATORS NEED TO EXAMINE THE CONTENT OF THE MUSIC PROGRAM NOT ONLY WITH REFERENCE TO THE KINDS OF ATTITUDES THIS PROGRAM SEEKS TO DEVELOP, BUT ALSO IN TERMS OF THE KINDS OF MUSICAL SKILLS AND COMPETENCIES WHICH GIVE MEANING AND PURPOSE TO SUCH A PROGRAM. (JL)

ED 002934

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.

DEVELOPMENT OF AUDITORY PERCEPTION OF MUSICAL SOUNDS BY CHILDREN IN THE FIRST SIX GRADES

Cooperative Research Project No. 766 (SAE 8411)

Dr. Robert George Petzold

University of Wisconsin
Madison, Wisconsin
June 30, 1960

The research reported herein was performed pursuant to a contract with the United States Office of Education, Department of Health, Education and Welfare.

3414a

CONTENTS

CHAPTER Page

I. INTRODUCTION.. 1

Introductory Statement
Need for Research Concerned with Auditory Perception
Purposes and Scope of the Study
Procedures and Hypotheses of the Study

II. SUMMARY OF RELEVANT RESEARCH............................ 8

III. PROCEDURES OF THE STUDY................................ 15

Analysis of Song Materials
Selections of the Sample
Description of the Sample
Construction of the Tests
Analyzing and Processing the Data

IV. RESULTS OF THE STUDY................................. 56

Introduction
Musical Experience
45-Item Test
Rhythm Test
Phrase Test
20-Item Test

V. SUMMARY AND CONCLUSIONS.............................. 103

VI. BIBLIOGRAPHY.. 112

VII. APPENDICES.. 113

A. Song-texts Analyzed for Tonal Configurations
B. Information Sheet; Scoring Key; Summary of Responses
C. Tests used in the Study; 60-Item Test, 45-Item Test, Rhythm Test, Phrase Test, 20-Item Test.

TABLES

Table Page

I. The Frequency Count of 267 Minor Tonal Configurations 17

II. The 64 Most Frequently Used Major and Minor Tonal Configurations 17

III. The 45 Most Frequently Used Rhythmic Patterns 22

IV. The Initial Selection of 602 Subjects, Showing Distribution by School and Grade 25

V. Final Distribution, by School, Grade and Sex, of the 606 Children Participating in the Study. 28

VI. Summary, by Grade, of Ages in Months for 467 Children 29

VII. Summary, by Grade, of Musical Training for 497 Children 29

VIII. Results of the Exploratory 60-Item Test 29

IX. The Rank Order of Difficulty and Discrimation for Items of the 60-Item Test 36

X. Reliability Coefficients and Standard Errors for the 45-Item Test 58

XI. Summary, by Grade, of Scores on the 45-Item Test 60

XII. Frequency Distribution, by Grade, of Scores on the 45-Item Test with References to Quartiles for the Total Sample 60

XIII. Summary, by Sex, of Scores on the 45-Item Test 63

XIV. Summary, by Experience Groups, of Scores on the 45-Item Test 65

XV. Summary of Types of Responses made by Children to the 45-Item Test 68

XVI. Summary, by Grade, of Scores for Children taking the Rhythm and 45-Item Tests. 72

XVII. Summary, by Sex, of Scores on the Rhythm Test 76

XVIII. Summary, by Experience Groups, of Scores on the Rhythm Test 77

TABLES (Continued)

Table		Page
XIX.	Summary, by Grade, of Scores on the Phrase Test	81
XX.	Frequency Distribution, in Percentages, of RL Scores on the Phrase Test	84
XXI.	Summary, by Sex, of Scores on the Phrase Test	88
XXII.	Summary, by Experience Groups, of Scores on the Phrase Test	89
XXIII.	Summary, by Grade, of Scores on the 20-Item Test	93
XXIV.	Summary, by Sex, of Scores on the 20-Item Test	96
XXV.	Summary, by Experience Groups, of Scores on the 20-Item Test	96
XXVI.	Summary of Responses to Items of the 20-Item Test	100

ACKNOWLEDGEMENTS

The writer wishes to express his thanks to the many Madison school children who participated in the study for their interest, enthusiasm and cooperation. He would also like to acknowledge the cooperation and assistance given to the project by Dr. Philip Falk, Superintendent of Schools, and Mr. LeRoy Klose, Supervisor of Music; as well as the many elementary school principals, classroom teachers, and special music teachers who provided facilities and made it possible to carry forward the extensive testing program. Recognition must also be given to the two research assistants, Mrs. Dorothy Williams and Mr. John Gates, who so competently fulfilled all responsibilities assigned to them.

CHAPTER I

INTRODUCTION

Introductory Statement

Music educators are agreed as to the significance of music reading skills in the musical development of the individual. Skill in music reading is considered an essential element of both musical understanding and appreciation, and of independent musical performance. Although there is little agreement regarding the degree of musical reading skill which children in the elementary and secondary school should be expected to develop, the basic music program of the schools gives considerable emphasis to a variety of activities designed to facilitate the development of a minimal level of such skill. There is also general agreement among music educators that a carefully planned "music reading readiness" program designed to provide opportunities for the child to develop an aural understanding of the melodic, rhythmic, and interpretative elements of music should precede the sequence of music reading activities. It is anticipated that the child will, as a result of growing competence in the aural perception of musical sounds, be able to proceed effectively to the visual perception of music symbols in music reading because they are already familiar with the "vocabulary" of music as represented by these sounds. Recommendations regarding the general nature of such "readiness" programs have been based upon collective practical experience rather than upon systematic research. Since the future musical growth of the child depends upon developing this aural understanding, it is necessary to be able to accurately identify the stages through which such development moves. It is particularily important that we fully understand the nature of this auditory perception when programs of instruction in music reading are planned.

The complex patterns of sounds which are called music have several characteristics which require the attention of the listener. Individual tones have pitch, duration, timbre and dynamic level or intensity. When individual tones are combined into groups, as long as there are a sufficient number of tones in the group to warrant consideration as a short musical entity, the characteristics of shape or contour, tempo, phrasing, tonality, and general style of performance become noticeable, as well as the harmonic implications of such a musical segment. The listener must, if he is to completely grasp the musical meaning of such combinations of tones, be capable accurately perceiving and reacting to all of the characteristics which may be present in the given musical situation, a task requiring simultaneous aural perception of these many characteristics. The writer is unaware of any significant research that has been carried forward in an attempt to determine the nature of aural perception with regard to the several elements present in music, or research which has been concerned with what happens, in terms of auditory perception, when these elements are presented in combination. A considerable amount of literature is available which discusses the relationship that the aural perception of certain basic musical elements bears to musical talent or capacity. In fact, many of the "standard" tests of musical ability include tasks which involve the aural perception of sounds, usually testing a single element at a time. They have not, however, provided the kinds of data necessary for careful study of the nature of auditory perception.

Need for Research Concerned with Auditory Perception

The primary need for an investigation of this kind arises from the lack of information available concerning the development of auditory perception by children in the elementary grades. Such information is necessary in order to plan the sequence and content of programs of instruction designed to develop an aural knowledge and understanding of the melodic, rhythmic and interpretative elements of music. Such aural understandings are considered necessary before the child can successfully begin to engage in activities which involve him with musical notation, the abstract symbols which are used to indicate the various characteristics of musical sounds. It is here assumed, and music educators appear to be in agreement with this assumption, that a knowledge of the "sound vocabulary" of music is of considerable value in learning the symbolism used to indicate those sounds. This is not to suggest that it would be possible for children to be exposed to the aural and visual aspects of music simultaneously - in fact, this is a widely accepted procedure followed in developing music reading skill. This procedure, however, requires some previous familiarity with the aural aspects of musical sounds in order for the visual symbolism to be fully understood by the child.

The writer became involved with this problem of auditory perception as he carried forward certain research relating to music reading by children in the upper elementary grades. The results of this exploratory investigation (10) revealed a relatively low level of music reading accomplishment and suggested that part of the difficulty might be attributed to lack of an aural understanding of musical sounds. A survey of professional literature failed to provide much information regarding the various levels of accomplishment associated with the auditory perception of musical sounds, or information concerning the developmental aspects of auditory perception. The present study was undertaken in an attempt to obtain such information.

Purposes and Scope of the Study

The basic purpose of the study was to determine the difference between children at each of the first six grade levels in the ways in which they perceive and respond to the auditory presentation of musical sounds. The identification of such differences would make possible a more precise definition of the nature of the development of the child's ear response to the melodic elements of music.

The study dealt primarily with the melodic element of music, that element later being visually perceived as those musical symbols which indicate pitch as the child begins to read music. The musical symbols used to indicate this melodic element of music are closely related to the aural impressions the listener has of the sounds - if the musical line moves up or down, so also do the written symbols on the printed page. This kind of relationship does not exist as we consider the rhythmic, or durative aspect of musical sounds.

Approximately 600 children, randomly selected from grades 1 through 6 of the total public elementary school population in Madison, Wisconsin were used in the study. The testing carried on in the study was all individual, with the data gathered by tape recording the pupil responses in the testing situations and later transforming these responses to syllabic notation entry on data sheets. Tape recording the pupil responses in the testing situation eliminated the need for immediate scoring at the time the test was given, thus insuring accuracy of the data since questionable responses could readily be repeated for verification.

One aspect of the study was concerned with the auditory perception of tonal configurations, with three tests developed to measure this type of response:

1. The aural presentation of 45 tonal configurations, varying as to length and complexity, presented in order of difficulty with the child singing the item he has just heard performed.
2. The aural presentation of these same 45 tonal configurations, with each configuration stated rhythmically as well as melodically, again requiring a singing response.
3. The aural presentation of 20 tonal configurations, with the total list randomized for each of three successive presentations, also requiring a singing response.

The first test provided data which would indicate possible levels of accomplishment since a considerable number of items of increasing difficulty were used. The second test provided data which made it possible to compare the performance on the aural perception of only the melodic element of music, as measured by the 45-Item Test just described, with performance on items which combined the melodic with the rhythmic. The third test gave data regarding the consistency of pupil response to a given stimulus.

A second aspect of the study was concerned with the auditory perception of more complete musical ideas as expressed by a musical phrase. Musical phrases, constructed from tonal configurations appearing most frequently in song materials, were aurally presented to the child, each presentation followed by a singing response. The child had ten trials in which to learn the phrase. This test provided data regarding the rate at which children were able to learn such musical phrases when they had to rely entirely upon the aural presentation of such material and did not have outside assistance in identifying the places where errors occurred.

For all of the tests, the subject was asked to make a singing response to the aural presentation of the item or the phrase. The vocal response was considered to be the most practical because all of the children had had experience in singing as a part of the regular music program carried on in the schools. Furthermore, there seemed to be no way of effectively measuring the accuracy of aural perception except in terms of some type of musical performance.

Procedures and Hypotheses of the Study

The procedures followed in the study are briefly outlined below and fully discussed in a later section of the report.

1. An analysis of song materials was carried forward to identify common tonal configurations which could be used as test items. A similar analysis identified common rhythmic patterns which could be used as test items.
2. The derivation of a stratified, random sample of grades 1 through 6; identification of the children within the sample; and subsequent collection of certain personal data for each subject constituted

the next step.

3. Construction of an exploratory test, followed by construction of the various tests used in the study.
4. The testing program was then carried forward. All children were given the initial 45-Item Test of tonal configurations. The total sample was then divided into three approximately equal groups, children assigned to the groups at random, and each group was then given one of the other three tests - 20-Item Test, the Phrase Test, or the Rhythm Test. This procedure made it possible to compare performance on the second test with performance on the first test.
5. Analysis of the data from the several tests, with comparisons being made between grade levels, musical experience levels, accomplishment levels, and between boys and girls.

The hypotheses of the study are as follows:

1. Age, musical experience and training are significant factors in the development of auditory perception.
2. The auditory perception of short tonal configurations precedes the auditory perception of larger musical units such as the musical phrase.
3. The development of auditory perception tends to parallel the development of vocal control.
4. Definite levels of accomplishment in terms of auditory perception may be identified.

CHAPTER II

SUMMARY OF RELEVANT RESEARCH

An extensive analysis of the professional literature relating to the musical development of children showed this literature to be grouped into two general categories:

a. Professional literature which includes courses of study, books in the field of music education which deal primarily with the methods of teaching music, and articles in professional journals.

b. The research literature in the fields of psychology, child development, music education and music.

This report will not concern itself with summarizing the information available in the first category because, despite its significance, these materials generally contain the opinions, personal experiences, and recommendations based upon personal experience which frequently serve as guides in planning programs of instruction in music education. Such information is readily available to the reader and need not be reviewed in a report of this kind. This literature shows that the authorities are in general agreement that one aspect of the musical development of the individual begins with developing, through those kinds of activities which involve the auditory perception of music, an awareness and understanding of the several elements of music. When the individual has achieved a reasonable degree of aural recognition of such elements it is believed that this establishes a satisfactory basis for involving the individual in those types of activities which give him experience in becoming familiar with the symbolic representation of the several musical elements. This completely logical sequence of events is not, unfortunately, a sequence which has been defined by careful investigation but has, despite this weakness, gained widespread acceptance.

The research literature dealing with the dual problems of auditory perception and the musical development of children is not extensive and this part of the report will concern itself with summarizing only those studies considered relevant to the present investigation.

Comparatively little information exists regarding the musical growth and the development of auditory perception for children of elementary school age. Several major studies dealing with the musical development of younger children, ages 2 1/2 to 6, point up the need for similar work with older children. The early work of Williams, Sievers and Hattwick (13), while concerned with pre-school children, represented a very promising beginning in the much-needed long-range investigation of the musical development of children. This series of studies described certain research techniques which the writer found to be of considerable value in planning the present study. The results of the various studies will not be given here because the investigators were not primarily concerned with children of elementary school age. They did develop a number of tests which have considerable merit in studying the general categories of sensory discrimination, complex cognitive abilities, motor control over the musical variables, and emotional responsiveness.

Drexler (2) worked with children, ages 3 to 6, in a study designed to obtain information regarding the ability of these children to carry a melody. The results of an extended training period, during which the children attempted to learn two simple songs, showed age to be a very significant factor in such achievement. The most significant differences were found between the 3-4 and 5-6 groups. Updegraf, Heiliger and Learned (12) found considerable individual differences in the ability to reproduce a tonal pattern by sings when the stimulus was presented aurally. Children, ages 3 to 5, participated in their study and were divided into two groups at each age level, matched in terms of musical ability as determined by pre-test scores and subjective teacher judgments regarding musical interest. One group at each age level was then given special music training for a specified period of time. Although the testing procedures were crude and the basis for analyzing and scoring the tests appeared highly subjective, the writer attributes some of these weaknesses to a lack of satisfactory equipment that now is available to the research worker. The general procedures were sound and the results of the study tend to support the hypothesis that age and musical training are significant factors in developing the kinds of competencies being studied. It was found that the five-year old child learned songs in approximately half the time required by four-year old children and that most of the three-year old children were able to learn only half of the material presented.

Jersild and Bienstock (5, 6) reported two studies dealing with the problem of developing singing ability in young children. The earlier study was concerned with the influence training had upon the number of tones and the number of intervals children could sing. They found, among other things, that training resulted in considerable improvement in extending the singing range and increasing the number of different intervals which could be sung. Their findings suggest that children in first and second grades who are identified as "uncertain singers" can benefit greatly from such training if it is available at these grade levels. The results of their second study showed that the range of tones children could sing increased rapidly from ages two to six, reaching a maximum number at approximately age eight, with this range then remaining relatively constant through adulthood. They also found that smaller musical intervals were more readily perceived and reproduced from aural stimuli than were the larger intervals of fourths and fifths. Jersild (7) was concerned with the vocal ability of children and raised the question as to whether those children who were given special training gain a permanent or temporary advantage over their peers who are without such training. If such training is only temporary then it would appear that maturation is a more important factor in developing vocal ability than is environment a conclusion which has considerable significance for planning programs of musical instruction. Unfortunately, the study was not carried on over the extended period of time necessary for reaching such conclusions but it did show that training resulted in a significant advantage over those without such training and that four months later the advantage was still apparent.

Hattwick (3) studied the average pitch range and pitch level preferred by children, ages 3 to 8, and found the level to be lower and the range narrower than was called for in the songs texts used in these first three grades. In fact, the same pitch level was preferred by all three groups and training seemed to have no effect upon raising the preferred pitch level so it more closely approximated the level called for in song texts.

These several studies are relevant to the present investigation because they emphasize the need for giving careful consideration to the pitch level and pitch range to be used in tests which require a singing response. They further indicate that vocal control is directly related to the concept of pitch for if good control is present, as evidenced by accurate singing of the aural stimuli, the concept of pitch has been differentiated from the other musical concepts present in the total musical experience. Since the writer planned to measure auditory perception in terms of a singing response to an aural stimulus it was necessary to establish the soundness of such a procedure.

Auditory perception is frequently defined as melodic or tonal memory and a limited, but significant body of research literature is available dealing with this subject. The studies of Drake, Karlin, Lundin, Kwalwasser and Dykema, Mainwaring, Dykema, Seashore, Semenoff and Wing have been concerned with developing tests capable of measuring musical talent, with musical memory as one significant factor in such talent. These reports will not be summarized here because, while the writer has examined these reports carefully, much of the information does not bear directly upon the present investigation. There are, however, certain studies which have a particular bearing on the present problem and these will be mentioned briefly.

Ortmann (9) defines the types of tonal patterns common in music and identifies the relative difficulty of each type. Subjects hear 5-tone patterns and then wrote what they heard, a test obviously designed for persons having considerable musical training and experience. He found that narrow intervals had a smaller range of error than wide intervals, that few subjects failed to recognize the element of repetition in any given pitch series, and also that few errors were made with regard to recognition of pitch direction.

Stepwise motion is easier to determine than motion by skips, disjunct pitch patterns which call for frequent directional change are the most difficult to recognize. In a more extensive article Ortmann discusses the psychological status of tones in melodic fragments and makes an analysis of typical test errors. (8) He considers the influence upon melodic memory of such factors as: (1) the length of the item, with the final tone of any pattern being stronger than the initial tone because it occurs last and is easily within the memory span, (2) the highest and lowest tones of a pattern tend to assume positions of prominence because they identify the pitch range of the item, and (3) the status of any given tone is not constant because it is associated with other factors present in a given melody. He also recognizes that tempo and intensity also affect the perception of groups of tones. Heinlein (4) was primarily concerned with questioning the validity of the Seashore Tonal Memory Test as one measure of a musical mind, particularily with reference to the nature of the task. The task required the listener to identify the changed tone within the second aural presentation of a tone pattern which made it different from the first presentation. Heinlein feels that the musical person perceives the total pattern as a complete entity and, for this reason, is unable to concentrate upon the separate parts of that pattern. He further points out that the number of tones in a pattern does not necessarily constitute a valid measure of the degree of difficulty, that some 6-tone patterns are easier than some 4-tone patterns. The Seashore Tonal Memory Test assumes that item difficulty is a function of the number of tones in the stimulus. Poly-direction patterns, says Heinlein, are generally more difficulty than uni-directional patterns and the factors of rhythmic grouping and accentuation functions in the perception of tonal configuration. Bugg and Herpel (1) were also interested in the nature of tonal memory as a function of musicality and chose to define "tonal memory" as the ability to remember isolated tones while "musical memory" was the ability to grasp and retain musical phrases. They concluded that tonal memory is an essential factor of superior pitch, timbre and rhythmic judgments.

This reported research emphasizes the need for studies which deal with the musical development of children of elementary school age. The literature fails to provide an adequate basis for recommending instructional procedures which will develop an aural understanding of the melodic elements of music that is necessary for understanding the visual symbolization of such elements. The data from the writer's previous study (10) showed a relatively consistent level of accurate response to aural stimuli by 4th, 5th and 6th grade children but there was little parallel success in responding to the visual presentation of these same stimuli as they read music. This inability to relate the visual symbols to the appropriate sounds appears to indicate that an aural understanding of musical elements is a necessary beginning as we consider the musical growth of children.

CHAPTER III

PROCEDURES OF THE STUDY

Analysis of Song Materials

One of the first steps taken in the present study was the analysis of basic song-texts to: (1) identify common tonal configurations; and (2) to identify common rhythmic patterns. The purpose of this analysis was to identify items which could be used in the construction of the various tests. Since the subjects for the study were to be selected from grades one through six it was necessary to have test items with which they would be reasonably familiar and, at the same time, have items which were capable of discriminating between the various levels of accomplishment. For this reason it was decided to construct test items on the basis of the frequency with which certain tonal configurations and rhythmic patterns appeared in the song materials used by children in these grade levels.

In a previous study (10) the writer carried forward an extensive analysis of song materials which identified common tonal configurations in songs written in major keys. The present study was to be concerned with the aural perception of minor tonal configurations, as well as major, and so the analysis was continued. The procedures followed in identifying tonal configurations have been detailed in this earlier report and will not be discussed here. The present analysis utilized 145 songs written in minor keys which were selected from seventeen basic song texts used in grades 2, 4, and 6. These songs were not selected at random but represent the total number of songs in minor keys appearing in these texts. It is quite evident that songs in major keys are more common than songs in minor keys, particularily since a detailed tabulation of 2515 songs showed that 2230, or 88.67 per cent of the total number of songs examined, were written in major keys with only 285, or 11.33 per cent of the total being written in minor keys.

Table I summarizes the numerical results of the analysis of minor songs. The total number of different minor tonal configurations identified in the 145 songs was 267, with a total frequency count of 1694 and a frequency range of 1 to 81. Examination of the data in Table I shows that the twenty tonal configurations having a frequency range of 21 to 90 constituted only 7.48 per cent of the total number identified but had a combined frequency count of 744, or 43.91 per cent of the total frequency count. Tonal configurations with a frequency range of 1 to 10 accounted for 224, or 83.9 per cent of the total number of configurations, but had a combined frequency count of 614, or 36.25 per cent of the total frequency count.

The previous analysis of 326 songs in major keys had given a total of 558 different tonal configurations with a total frequency count of 4575 and a frequency range of 1 to 213. It had been noted in this earlier analysis that 22 configurations, or not quite 4 per cent of the total number, with a frequency range of 44 to 213 had a combined frequency count of 1939, or 42.38 per cent of the total frequency count. In this same analysis, those configurations with a frequency range of 1 to 9 accounted for 82.44 per cent of the total number of configurations identified, but only 29 per cent of the total frequency count. Both sets of data point up the fact that there exists a relatively small number of tonal configurations which appear repeatedly in

Table I

The Frequency Count of 267 Minor Tonal Configurations

Frequency Range	Number of Configurations	Per cent of Total Count	Combined Frequency	Per cent of Total Frequency
81 - 90	1	0.37	84	4.95
71 - 80	0	0	0	0
61 - 70	1	0.37	70	4.13
51 - 60	0	0	0	0
41 - 50	5	1.87	241	14.23
31 - 40	4	1.50	130	7.67
21 - 30	9	3.37	219	12.93
11 - 20	23	8.61	336	19.83
1 - 10	224	83.90	614	36.25
Totals	267	99.99	1694	99.99

Table II

The 64 Most Frequency Used Major and Minor Tonal Configurations

Configuration	Frequency	% Total Frequency	Configuration	Frequency	% Total Frequency
so do	213	3.4	so mi do	93	1.5
mi re do	198	3.2	so mi	90	1.4
do re mi	155	2.5	do ti do	87	1.4
so la so	138	2.2	la ti do re mi	86	1.4
mi fa so	115	1.8	so la ti do	80	1.3
do ti la so	111	1.8	do ti la	77	1.2
do mi so	96	1.5	mi so	76	1.2
so fa mi	95	1.5	la ti do	73	1.2

Table II (Continued)

Configuration	Frequency	% Total Frequency	Configuration	Frequency	% Total Frequency
mi la	73	1.2	do mi do	33	0.5
do la so	72	1.2	do re mi do	31	0.5
do so	68	1.1	fa mi re	30	0.5
so mi so	62	1.0	do re mi re do	30	0.5
mi fa mi	61	1.0	mi re do re mi	30	0.5
so fa mi re do	60	1.0	do do	29	0.5
fa mi re do	58	0.9	so fa mi re do	26	0.4
do mi	55	0.9	do so do	26	0.4
mi do	55	0.9	so do	26	0.4
mi re mi	54	0.9	la mi	26	0.4
mi re do ti la	53	0.9	so mi	25	0.4
do re do	50	0.8	mi re do la	25	0.4
la si la	47	0.8	mi do ti la	25	0.4
so do mi	44	0.7	fa re	24	0.4
re mi fa	44	0.7	so la so fa mi	23	0.4
do re mi fa so	41	0.6	la ti do ti la	23	0.4
la so fa mi	40	0.6	mi so do	22	0.3
re mi re do	39	0.6	mi re do ti	22	0.3
do la	38	0.6	la mi	22	0.3
mi do la	34	0.5	so do so	22	0.3
la so mi	34	0.5	mi la	21	0.3
do la do	33	0.5	re fa mi re do	21	0.3
mi so mi	33	0.5	mi do re mi	21	0.3
mi la mi	33	0.5	so fa mi re	20	0.3

the songs children sing.

The construction of test items was to be based, in part, upon the frequency with which certain tonal configurations appeared in songs and so the data from the two analyses were combined. This gave 64 common major and minor tonal configurations having a frequency range of 20 to 213. These 64 configurations, or 7.76 per cent of the total number identified in both analyses, accounted for 57 per cent of the total frequency count and are listed in Table II, page 18. These 64 configurations were then used in the construction of the melodic test items for the several tests.

Although the study was to be primarily concerned with the auditory perception of the melodic elements of music, the writer felt that it would be of some interest to briefly explore the effect upon such perception when the rhythmic element was combined with the melodic element. This decision made necessary the analysis of song materials to identify common rhythmic patterns. A detailed tabulation of 2515 songs found in the basic song-texts used in the elementary grades showed that 1530, or 60.83 per cent of the total number of songs utilized time signatures of $\frac{4}{4}$, $\frac{2}{4}$, or $\frac{2}{2}$ and that 931, or 37.02 per cent utilized $\frac{6}{8}$, $\frac{3}{4}$, or $\frac{3}{8}$ time signatures. Songs written in these six time signatures accounted for 2461, or 97.85 per cent of the total number of songs examined. It is difficult for the musically untrained individual to aurally differentiate between certain basic rhythmic patterns found in the various duple meters or between those in the various triple meters, because such differences are largely notational in character rather than differences in sound. For example, assuming that no marked differences in tempo exist, the 𝅗𝅥 ♩ ♩ pattern in a $\frac{4}{4}$ measure will sound much like the ♩ ♫ pattern in a $\frac{2}{4}$ measure. Similarily, the 𝅗𝅥 ♩ pattern in $\frac{3}{4}$ sounds very much like the ♩♪

pattern in $\genfrac{}{}{0pt}{}{6}{8}$ measure. Because of this lack of difference, in terms of sound, the writer decided to limit his analysis to those songs written in $\genfrac{}{}{0pt}{}{3}{4}$ and $\genfrac{}{}{0pt}{}{4}{4}$ in order to avoid unnecessary duplication of effort. A total of 1047 songs, selected from fifteen basic song-texts used in grades 2, 4, and 6 were examined. Of this total, 661 songs were in $\genfrac{}{}{0pt}{}{4}{4}$ and 386 in $\genfrac{}{}{0pt}{}{3}{4}$. These books are listed in Appendix A.

The major problem of establishing objective criteria for accurately identifying rhythmic patterns still remains to be solved. The writer was unable to discover any research which proved helpful in this regard despite the fact that developing the ability to accurately read rhythmic patterns represents a significant problem area for the music teacher, particularily the instrumental music teacher. One factor which makes difficult the consistent application of any criteron for identifying rhythmic patterns in songs is the interrelationship which exists between the words of the song and the rhythmic notation of the melodic line. For example, the rhythmic notation of any given song depends upon the number of syllables in the words and, as the words may change from verse to verse, so also should the rhythmic notation change. Notational convention, combined with printing economy, usually fails to indicate such changes satisfactorily and the individual singer is forced to adapt the text to the melodic line as effectively as his musical training permits.

The writer arbitrarily established the musical measure as the basis for identifying the length of a rhythmic pattern and, in certain instances, the half-measure became the more logical basis for such identification. In situations where neither of these seemed to apply, the text which accompanied the melodic line was examined and the decision then reached. In general, the rhythmic notation for only the first verse of any song was analyzed. A total of 99 different rhythmic patters were identified, with a total frequency count of 9111 and a frequency range of 1 to 1175. Of this total, 58 different rhythmic patterns appeared in songs having $\frac{4}{4}$ time signatures and 41 in songs with $\frac{3}{4}$ time signatures. The 45 most common rhythmic patterns are listed listed in Table III, pages 22-23. The data summarized in this table shows that the 20 patterns in $\frac{3}{4}$ time have a total frequency count of 3940, or 94.4 per cent of the total frequency count for all patterns in this meter. It is also significant that the 25 patterns in $\frac{4}{4}$ account for 43-2, or 87.3 per cent of the total frequency count for patterns in this meter. These 45 patterns were considered a sufficient number from which test items could be drawn and there was no need to consider the remaining 54 patterns which had been identified.

Selection of the Sample

The design of this study called for a stratified sample from the normal population of children from grades one through six of the Madison, Wisconsin public elementary schools, this population being undifferentiated in terms of musical ability. There are 24 public elementary schools in the city of Madison but the representative sample of approximately 600 pupils was randomly selected from only 20 of these schools. The pupils in two of the 24 schools were not available for testing and pupils in two other schools had participated in an exploratory study connected with the project and so these four schools were excluded from the study. It had been determined that approsimately 600 pupils could be tested during the period of time available for

TABLE III

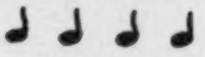
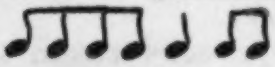
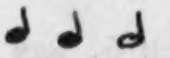
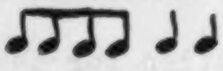
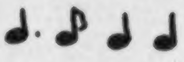
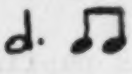
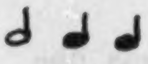
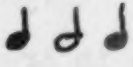
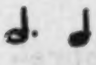
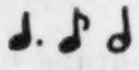
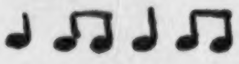
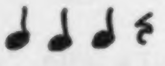
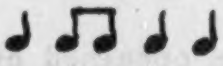
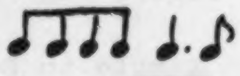
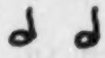
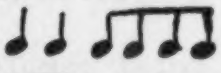
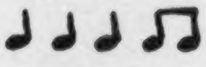
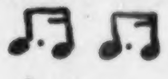
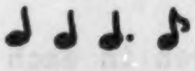
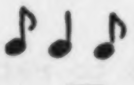
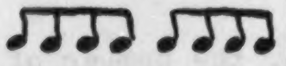
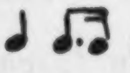
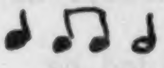
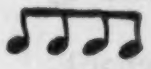
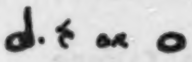
The 45 Most Frequently Used Rhythmic Patterns $\binom{3}{4}$

Pattern	Frequency*	Per Cent**	Pattern	Frequency	Per Cent
♩ ♩ ♩	1175	28	♫ ♫ ♫	70	1.7
𝅗𝅥 ♩	710	17	♩ 𝄽 ♩	60	1.4
𝅗𝅥.	366	8.7	♩ ♩ 𝄽	46	1.1
♩. ♪ ♩	233	6.8	♫ ♩ ♫	41	1.0
𝅗𝅥 ♫	177	4.2	♫ 𝅗𝅥	36	0.9
♩ ♩. ♪	148	3.5	♩ ♫ ♩	29	0.7
♩. ♪ ♫	144	3.5	♩ ♩ ♪.𝅘𝅥𝅯	27	0.7
♩ ♫ ♫	137	3.3	♩ ♪.𝅘𝅥𝅯 ♪.𝅘𝅥𝅯	27	0.7
♫ ♩ ♩	137	3.3	Total	3940	94.4
♩ ♩ ♫	129	3.1			
♩ 𝅗𝅥	117	2.8			
𝅗𝅥 𝄽	81	1.9			

* Frequency count in $\genfrac{}{}{0pt}{}{3}{4}$

** % of total frequency count in $\genfrac{}{}{0pt}{}{3}{4}$

TABLE III (Continued)
The 45 Most Frequently Used Rhythmic Patterns
$(\frac{4}{4})$

Pattern	Frequency*	Per Cent**	Pattern	Frequency	Per Cent
	797	16.2		89	1.8
	367	7.4		89	1.8
	361	7.3		65	1.3
	212	4.3		58	1.2
	212	4.3		57	1.2
	208	4.2		53	1.1
	207	4.2		57	1.2
	202	4.1		55	1.1
	198	4.0	 (½)	150	3.0
	169	3.4	 (½)	106	2.2
	114	2.3	 (½)	103	2.1
	109	2.2			
	108	2.2	Total	4302	87.3
	156	3.2			

* Frequency count in $\frac{4}{4}$

** % of total frequency count in $\frac{4}{4}$

such purposes. This stratified sample, with the pupils selected at random within each grade level, maintained the appropriate proportions of the total population in terms of school size, enrollments in a given grade, and number of boys and girls.

The following steps were taken in deriving this sample:

1. The total population for these 20 schools, grades 1 through 6, was obtained, giving 8432 pupils.
2. The number of pupils in grades 1 through 6 in each school was obtained and the percentage of the total population which was represented by this figure was calculated for each school.
3. The size of the sample to be drawn from each school was obtained by multiplying the total sample of 600 by the percentage obtained in Step 2.
4. Since the sample was to retain that proportion of children in each grade level which was found in the total population, the number of children in each of the six grades in each school was divided by the total enrollment of the school. The sample for each school, obtained in Step 3, was then multiplied by the percentage figure for each grade to give the number of children for the sample from each grade in each school.
5. The total school population contained approximately equal numbers of boys and girls and this proportion was retained, as nearly as as possible, within the sample.

Table IV, page 25 lists the number of pupils which were to be selected from each grade in each of the 20 schools.

Table IV

The Initial Selection of 602 Subjects, Showing Distribution by School and Grade

School	N	% Total N	n	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6
1	166	1.96	12	2	2	2	2	2	2
2	314	3.73	22	4	4	4	4	3	3
3	592	7.02	42	9	7	6	6	6	8
4	397	4.71	29	5	5	5	4	5	5
5	303	3.59	23	6	4	3	4	3	3
6	399	4.73	27	5	4	5	5	4	4
7	206	2.45	15	3	3	2	2	3	2
8	243	2.88	18	4	3	3	3	3	2
9	497	5.89	33	9	5	4	5	5	5
10	534	6.33	39	9	6	6	6	6	6
11	405	4.81	28	6	6	4	4	4	4
12	594	7.04	44	8	7	7	7	8	7
13	349	4.14	25	4	4	5	4	4	4
14	200	2.37	16	3	4	2	3	2	2
15	356	4.22	26	7	5	4	4	3	3
16	451	5.35	32	5	5	6	5	5	6
17	1011	11.99	70	14	14	13	12	10	7
18	794	9.41	55	12	11	7	8	8	9
19	344	4.08	26	5	4	5	4	4	4
20	277	3.28	20	4	4	3	3	3	3
Total	8432	100.00	602	124	107	96	95	91	89

N = Enrollment in each school, grades 1-6.
n = Size of sample

The procedures followed in identifying specific subjects within the sample were as follows.

1. The class lists for each grade in each school were obtained. The pupils were then listed alphabetically, by sex, within each grade and assigned consecutive numbers. Each grade, therefore, had two lists, one of boys and one of girls, with the pupils in each list numbered from 1 to "i".
2. A table of random numbers was used, with the highest number to be drawn from the table corresponding to the highest "i" on either list for the grade in question. A single number drawn from the table identified both a boy and a girl because of the dual lists. For example, if the number 12 was drawn, the twelfth boy and twelfth girl were then identified for testing. Sufficient numbers of subjects were randomly drawn from each grade to give the totals listed in Table IV, plus alternates in the event a pupil was not available for testing.

The testing program began as soon as all subjects had been identified. The schedules for the music classes had been obtained from each school and an elaborate time table for testing had been constructed.

Description of the Sample

A total of 660 children were given the basic 45-Item Test of tonal configurations during the project, the additional children serving as alternates in the event some children were unavailable at the time the second test was to be given. Of this total, usable results were obtained for 606 children, distributed by school, grade and sex as shown in Table V, page 28. It proved impossible to maintain the original distribution, as presented in Table IV, page 25, because children were absent from school, others had moved to other communities or other schools within the city, the music classes were cancelled because of other school activities, and other similar situations. Furthermore, certain test data was not usable because some of the children did not complete the basic 45-Item Test. The reasons for failure to complete the test were of two kinds: (1) the subject verbally expressed no desire to continue the tests, or (2) the subject became so nervous during the testing that the examiner considered it wise to stop, since continuing the test might not yield a true measure of the child's ability to perform the task.

An Information Sheet, Appendix B, designed to gather certain types of personal data, was given to each child who participated in the study. Despite several follow-up contacts, there were many children who failed to return the sheet or who did not respond to all items on the sheet. The information from these sheets showed that although the range of ages at each grade level was considerable there was little variation in age within a given grade. This data lent support to the writer's decision to compare groups in terms of grade level rather than in terms of specific age in months. Table VI gives the average age, range, and standard deviation for the children in each grade for whom data was available.

The Information Sheet sought data regarding the musical training of the children as defined in terms of playing and studying a musical instrument. Responses were secured from 497 of the 606 children and Table VII, page 29 summarizes these responses. The data shows the anticipated marked increase

Table V

Final Distribution, by School, Grade and Sex, of 606 Children Participating in the Study

School	Grade 1		Grade 2		Grade 3		Grade 4		Grade 5		Grade 6	
	B	G	B	G	B	G	B	G	B	G	B	G
1	2			1	2		1	1				
2	2	2	3	4	3	1	4	4		3	4	
3	9	3	1	5	3	3	3	2	1	3	2	5
4	2	2	2	2	5	1	3	1		4	3	2
5	2	4	2	1	2	2	3	2	1	1	2	
6	4	1	3	1	1	3	1	5	1	3	3	1
7	1	2	2	2	1	1	1	1	1	2		3
8	1	2	1	2	3	1	2	2	2	2	2	
9	1	7	3	1	1	2	1	4	3	5	3	3
10	4	3	1	3	1	2	4	3	4	3	1	1
11	3	2	1	4	3	1	1	3	2	2	3	1
12	6	3	2	4	4	4	4	3	7	3	4	2
13	2	2	3	3	4	3	3	3	4	4	6	2
14	1	1	3	1	1	1	2	1		3	2	1
15	1	5	1	3	1	3	1	2	2		2	
16	2	2	3	4	1	1	2	4	5	4	3	4
17	3	4	3	9	6	5	6	6	4	8	2	3
18	5	8	5	6	5	2	2	5	2	6	5	3
19	8	4	3	1	3	3	2	2	2	4	4	4
20	3	1	1	3	1	3	1	1	2		2	
Total	62	58	43	60	51	42	47	55	42	60	52	34
Total	120		103		93		102		102		86	

Table VI

Summary, By Grade, of Ages in Months for 467 Children

Grade	N	Sum of X	Sum of X^2	Mean	S. D.	Range
1	86	7170	600,898	83.4	6.02	31
2	84	8115	786,143	96.6	5.09	24
3	68	7341	795,361	107.9	6.50	33
4	81	9676	1,157,628	119.5	4.68	29
5	83	11001	1,459,973	132.5	4.76	34
6	65	9308	1,334,376	143.2	4.75	29

Table VII

Summary, By Grade, of Musical Training for 497 Children

Grade	N	n who play	% Total N
1	94	0	0
2	91	9	10.0
3	71	11	15.5
4	84	17	20.2
5	90	42	46.7
6	67	28	41.8
Total	497	107	

Table VIII

Results of the Exploratory 60-Item Test

	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 1-6
N	14	21	13	24	27	24	123
Score Range	1-49	2-50	7-46	3-55	0-59	1-57	0-59
Mean	23.9	25.0	25.2	32.0	29.4	37.8	29.7

in the number of pupils studying a musical instrument from grade level to grade level, with the exception of the sixth grade. The figures for the fifth and sixth grades reflect the influence of the instrumental program in the elementary schools which provides instruction for all interested children. Further analysis of these responses showed that of the number of pupils indicating that they played a musical instrument, 23 had studied less than one year, 44 had studied at least one year, 22 reported having studied for two years, 10 had studied for at least three years, and 8 had studied their instrument for four or more years.

The data regarding the musical activity in the home which was obtained from the Information Sheets will be discussed later in the report.

Construction of the Tests

The purposes of this study were; (1) to examine perception of musical sounds as the aural identification and recognition of the similarities and differences between tonal configurations by children in the first six grades; (2) to study the techniques appropriate for the collection and analysis of such data; and (3) to provide data which could be used to suggest research procedures that could be applied in subsequent studies of this problem area. Two types of tests were needed: (1) a test which would give data regarding the aural perception of short tonal configurations, varied as to length and complexity; and (2) a test which would give data regarding the auditory perception of larger and more complete musical ideas found in the musical phrase, both in terms of the rate at which children learned such phrases as well as the amount of recognition taking place.

The analysis of song materials discussed earlier in the report had resulted in identifying common tonal configurations and rhythmic patterns which could be used as possible test items. The time element was an important factor in developing the final form of each test since arrangements to use children from the schools had been approved with the limitation that all such testing was to be done during the child's regular music period. It had been discovered in an earlier study that children were unable to work effectively and efficiently on tasks requiring considerable concentration for periods of time in excess of twelve to fifteen minutes. A third reason for constructing tests which could be completed within a ten- to fifteen-minute testing session was the necessity for considering the scheduled time given to the music classes in order to use this time efficiently. The music periods in the Madison schools range from 20 to 45 minutes in length, varying somewhat from school to school as to length of period and number of periods per week. In view of the number of pupils who were to be tested during the study it was necessary to construct tests of such a length so as to permit testing more than a single pupil during a given music period. A ten-minute test would satisfy such a condition.

1. The Exploratory Test

Certain preliminary exploration was necessary before the basic test could be constructed. It was necessary to develop a routine for testing, as well as to select test items which would discriminate between various levels of accomplishment and provide a range of item difficulty. An exploratory test of 60 items of varying length and complexity was developed. The construction of the test items utilized some of the suggestions of Ortmann (8, 9), who was interested in defining and measuring tonal memory in music, as well as data gathered in the writer's own previous research. (10) This 60-Item Test contained equal numbers of 3-, 4-, 5-, 6-, and 7-tone patterns, representing major and minor ascending and descending scale and chord figures. A copy of this test is included in Appendix C.

The sixty items were arranged in five sets of twelve, with each set becoming progressively more difficult than the preceding set. The writer was forced to make subjective judgments regarding the difficulty of an item although he was guided, in the case of many items, by an item analysis carried forward previously. (10) To insure uniformity of testing procedures, the aural presentation of test items played by a piano, was tape-recorded at a tape speed of 3.75 inches per second. An interval of silence, during which the subject was to sing what had been played, followed each aural presentation with this time interval varying according to the length of the item. The timing was managed by setting an electric metronome to mark every second, using the flashing light rather than clicks to keep the test tape free of unnecessary sounds. The test tape contained the following: (1) the number of each item was announced, (2) this was followed by the aural presentation of the item played at a rate of two quarter notes per second, (3) followed in turn by the interval of silence. The same sequence was repeated for each of the remaining items. The durations of these intervals of silence were as follows: 4 seconds for 3- and 4-tone items, 5 seconds for 5- and 6-tone items, and 6 seconds for 7-tone items. This allowed sufficient time for response and the total test time was ten minutes. A brief orientation to acquaint the subject with the nature of the task was developed and included five sample test items which were placed at the beginning of the test tape. The total time for any testing session seldom exceeded fifteen minutes and most subjects were able to complete the test in twelve minutes.

Each child was tested individually throughout the exploratory study, as well as the major portion of the project. Equipment used in the test situation included two tape recorders (Wollensak, Model T-1500), one for playing the test tape containing the aural stimuli and the second for recording all pupil responses. Usable data for this 60-Item Test was obtained from 123 children selected at random from grades one through six in two of the elementary schools, also selected at random from the total number of schools. Since the basic purpose of this exploratory testing was to provide data needed for an item analysis, the results of the test will not be discussed in the report. Table VIII, page 29 gives the number of pupils tested, together with the mean score for each grade and the range of scores.

The 60-Item Test was scored on a "right" or"wrong" basis with the highest possible score being 60, one point for each item correctly reproduced. The exploratory testing revealed no significant weaknesses in either the test itself or in the procedures which were followed. The subjects reported that the stimulus (piano) was easy to hear and that they had sufficient time to sing a response. Many of them did report having difficulty in remembering the starting pitch of the item but the writer felt that any attempt to give such a pitch immediately after the aural presentation of the item would serve to confuse most of the children and might make them forget the item completely. It was felt necessary to reduce the content of the test so it more closely approximated a ten-minute testing session and to refine the scoring procedure so that greater flexibility was possible. The matter of revised scoring will be discussed later in the report. A more detailed set of instructions, to be followed by the examiners, was developed to insure minimum variation from test situation to test situation.

The item analysis of the test followed the procedures suggested by Ross and Stanley (11) and are outlined below:

1. Since all subsequent tests were to be administered to children from grades one through six, the six grades were combined in carrying forward the item analysis.
2. Two groups were then identified: the 27 per cent making the highest test scores [W_H]; and the 27 per cent making the lowest test scores [W_L]. There were 33 subjects in each group (n).
3. Item difficulty was computed by summing the number of "wrong" responses made by pupils in both groups for each item [$W_L + W_H$] and multiplying by a constant $\left[\frac{100 \text{ times the number of options}}{2\text{ n (number of options)}}\right]$ to correct for chance. The resultant figure was the measure of difficulty for the item. The larger the number the more difficult was the item. The items were then ranked according to difficulty, from "easy" to difficult".
4. Item discrimination was computed by subtracting the "wrong" responses of the upper group from the "wrong responses of the lower group [$W_L - W_H$], the larger this figure is the more discriminating power the item has. The 60 items were then ranked in order of

discriminating power from "poorest" to "best".

Table IX, page 36 summarizes the results of the item analysis. Since many of the items had identical values for either item difficulty or item discrimination, they were ranked by item number, identical rank scores indicating such ties.

2. 45-Item Test

The design of the study called for a basic test which would provide data regarding the auditory perception of short tonal configurations. Such data would make it possible the identification of various levels of accomplishment as well as to confirm the existence of possible growth patterns in the development of auditory perception. To accomplish this it was necessary that the test should contain a reasonably large number of items which varied in difficulty and complexity.

Items for this basic 45-Item Test were selected on the basis of the item analysis of the 60-Item Test. Items were elini ated which ranked either "very easy" or "very difficulty" and which also ranked "poor" in discriminating power. An attempt was made to retain almost equal numbers of 3-, 4-, 5-, 6- and 7-tone patterns, as well as to insure that the patterns themselves represented the following kinds of musical contours: (1) ascending scale pattern, (2) ascending chord pattern, (3) descending scale pattern, (4) decending chord pattern, (5) ascending and descending scale pattern, (6) ascending and descending chord pattern, (7) disjunct patterns, and (8) combinations of scale with chord, both conjunct and disjunct.

The final form of this basic test included 45 such items since this was the number which would permit completion of the test within a ten to twelve

Table IX

Rank Order of Difficulty and Discrimination for Items of the 60-Item Test

Item No.	Diff. Rank	Discrim. Rank	Item No.	Diff. Rank	Discrim. Rank	Item No.	Diff. Rank	Discrim. Rank
1	3.5	10	*21	38.5	51	**41	32.5	58.5
*2	22.5	48.5	**22	32.5	54	*42	46	29
3	7	22.5	23	11	16	**43	40.5	39
*4	11	22.5	*24	20	35.5	**44	44	39
*5	22.5	54	*25	17	35.5	*45	46	16
*6	9	26	**26	36	48.5	**46	35	56
**7	29	44	27	3.5	10	*47	20	44
**8	13.5	32	**28	32.5	54	*48	40.5	26
9	5	16	**29	38.5	51	*49	50	19.5
**10	25	44	**30	25	44	*50	53	10
*11	20	35.5	**31	46	29	*51	48.5	26
*12	15	35.5	*32	7	22.5	**52	42.5	44
13	1	10	*33	13.5	32	*53	51	16
14	7	22.5	*34	18	39	54	58.5	3
*15	16	32	*35	53	10	**55	42.5	44
**16	29	44	**36	29	51	*56	48.5	19.5
17	2	13	**37	27	58.5	**57	37	58.5
*18	25	29	38	53	7	58	56.5	4.5
**19	32.5	58.5	39	56.5	4.5	59	60	1
*20	11	16	40	58.5	2	60	55	6

Difficulty Rank-The larger the number, the more difficult is the item.
Discrimination Rank-The larger the number the poorer is the discrimination of the item.

*** Items used in 45-Item test. **Items used in 45- and 20-Item Tests.**

minute session. The items were arranged in order of item difficulty, from "easy" to "difficult", and a copy of this test is included in Appendix C. The reader is referred to Table IX, page 36 giving the data resulting from the item analysis where the items selected for the basic test are identified by "*". The procedures for making the test tape and for administering the test were identical with those followed for the 60-Item Test. The pupil was first given a brief orientation by the examiner to acquaint him with the task. This was followed by a trial of five sample items presented at the beginning of the test tape. The test tape was then stopped to permit the pupil to raise questions as well as to give the examiner an opportunity to clarify the requirements of the task and put the pupil more at ease if necessary. Following this, the test tape was again started, as well as the second tape recorder which recorded the complete test and the pupil's response to each item. The details regarding the processing of the test data will be discussed later.

3. 20-Item Test

A second test used in the study was designed to give data regarding the consistency of pupil responses to a given stimulus. For this purpose the writer decided to use three presentations of a stimulus and, still keeping in mind the time factor, this would permit using 20 to 25 items, the total number depending upon the length of the items selected.

Reference to the item analysis of the 60-Item Test resulted in identiyying 30 items having high discriminating power which also ranked in the middle of the difficulty range. From this list 20 items were selected for the test, with the following criteria serving as a basis for selection:

1. Items would be selected primarily on the basis of discriminating power.
2. In cases where items with the same discriminating power were chosen, the preference in making the final selection would favor items which were also difficult rather than easy.
3. Equal numbers of 3-, 4-, 5-, 6-, and 7-tone items were to be used.

The twenty items were then randomized for each of three successive trials and the test placed on tape according to the procedures already outlined for the 60-Item and 45-Item Tests. The items were randomized for each trial because the purpose of the test was to ascertain the consistency of response rather than to identify the rate at which children were able to learn the item. Had three successive presentations of the same item been made before going on to the next item it is doubtful whether the data could have been interpreted in terms of consistency of response since the element of rote learning would affect the score. A copy of the 20-Item Test is included in Appendix C.

4. Phrase Test

The tests constructed thus far were designed to give data regarding the auditory perception of short tonal configurations. Music, although it might be reduced to such small melodic segments for purposes of analysis, is generally considered to involve larger, more complete musical ideas. The musical ideas have definite melodic shapes or contours; they usually will have a recognizable tonality, especially the kinds of songs children are likely to sing; and there is almost always movement from one point of repose, through one or more points of tension, toward another point of repose. Much of the music learned by children in the elementary grades is the result of extensive rote teaching, a procedure which appears to be retained in teaching vocal music at even higher levels of our educational system. This means that if children are to learn songs rapidly and effectively they need to develop a sense of tonal memory which enables them to readily perceive and retain the essential elements of such larger musical ideas as the musical phrase. It is particularily important that they be able to retain the sound of the phrase long enough so they are able to judge the accuracy of their own performance in terms of the original stimulus.

The writer has long felt that the auditory perception of larger musical units represents a highly significant problem area in the musical growth and development of the individual. For this reason it was decided to include in the present study a short test concerned with this problem. The intent of such a test was not to enable the writer to reach any definite conclusions but rather to offer data which could suggest research techniques and procedures which might be employed to effectively approach the entire problem with greater objectivity.

In constructing such a test the following guides were utilized:

1. The musical unit, preferably a four-measure phrase, should be reasonably musical without duplicating any familiar song.
2. Each unit, or phrase, should utilize only the most common tonal configurations identified in the analysis of song materials so most subjects would be capable of learning at least part of the

phrase.

3. The melodic line should be written so as to include a simple, repetitive rhythmic pattern to permit easier recognition, more consistent organization, and greater retention since there were to be no accompanying words which could be associated with the melody. Although it has not been established by systematic investigation, it is usually agreed that children are able to learn songs more readily if they associate the words of the song with the melodic line. In the absence of words, the rhythmic structure of this melodic line appears to provide the child with an easily learned "cue" which simplifies the learning of the melodic line. The rhythmic structure tends to organize the melodic line into shorter more recognizable units.
4. The phrase should encompass a reasonable singing range so as to permit the children to sing easily and without tension. Children frequently alter the melodic line of the song they are singing, not because they are unable to retain the image of this melody but be- because such a change makes it possible to keep the song within because such a change makes it possible to keep the song within their comfortable singing range.

A number of different phrases were constructed in both major and minor modes, and submitted to an informal jury of composition students, theory teachers and elementary school music teachers. The jury was asked to identify those phrases which seemed most musical as well as to make judgments regarding their relative difficulty. Two phrases, one in major and one in minor, were finally selected for use in the test. These phrases, with the tonal configurations identified, are included in Appendix C. The singing range of both phrases went no lower than B♭ and no higher than a C♮ a ninth higher and so it could be assumed that children would encounter little physical difficulty in singing. Each of these phrases utilized a simple rhythmic pattern which was repeated throughout the four measures. This rhythmic pattern made the task of scoring much easier because it was possible to accurately identify the measure, or part of a measure, which the subject was singing.

Each phrase was played at a tempo of 𝅗𝅥. = 60 and ten presentations of each phrase were recorded on the test tape. The subject was told that he would hear a short song twice and then he was to try to sing as much of it as he could remember. Following this first singing trial the subject heard the phrase played a third time, this hearing again followed by a singing trial. This process of listening and singing continued until the subject had nine opportunities to sing the phrase or until he had sung it perfectly two trials in succession. This criterion of success was used to eliminate the necessity for continuing with the test when the subject had demonstrated that no more learning could take place. The examiner gave the subject the starting pitch of the phrase immediately prior to each singing trial because several exploratory testing sessions had demonstrated that it was not possible to assume that children could locate this pitch without such help. All pupil responses were tape-recorded for later analysis and processing.

The total time for each phrase, provided a subject heard all ten presentations, was ten minutes. Since the total test time available was not to exceed twelve minutes for reasons given earlier, this meant that children would be unable to proceed to the second phrase unless they had learned the first phrase in five or six trials. It was considered inadvisable to prolong the testing session to permit the presentation of the second phrase and the duration of the project did not permit scheduling additional testing sessions with these same subjects.

The procedures for processing and scoring the phrase test will be discussed at another point in the report.

5. Rhythm Test

The tests described thus far have attempted to measure the auditory perception of only the melodic element of music. Except for the Phrase Test, all other tests utilized tones having equal duration. Most of the music we hear and perform represents a combination of the melodic and rhythmic elements, not the presentation of a single element. The rhythmic structure of a melodic line helps give that melody an identity which is unique, a factor which is of considerable assistance in learning and retaining that melody.

The writer is interested in exploring the interrelationships which may exist between these two basic elements of music in terms of their effect upon the auditory perception of musical sounds. The scope of the present study would not permit a detailed examination of the problem but it would allow certain preliminary investigation which might suggest research techniques and procedures for later use. It was decided, therefore, to administer the 45-Item Test to one group of subjects with the items re-written so that each represented a common rhythmic pattern as well as a common tonal pattern. The analysis of song materials had identified 45 such common rhythmic patterns and these were used as a b sis for re-constructing the test items. The following criteria served as guides in determining that rhythmic pattern which would be most appropriate for a given tonal configuration:

1. The number of tones should be the same. For obvious reasons five-tone rhythmic pattern could not be superimposed upon a four-tone tonal pattern without duplicating some tone within the pattern and this was not to be permitted.
2. The movement characteristics of the rhythmic pattern should, as far as possible, coincide with the movement characteristics of the tonal pattern. This is a difficult criterion to verbalize but means that if the tonal pattern ends with a full or partial cadence, thus establishing a point of repose, the rhythmic pattern should do the same thing. Item 5 of the test is an example of this; the descending tonic chord in the minor comes to rest on the tonic tone, the rhythmic pattern also tends to come to rest on its third tone.
3. Many different rhythmic patterns should be incorporated in the test rather than duplicating a limited number. This would seem to insure careful listening to the entire item, in terms of both the melodic and rhythmic elements, since it would not permit the attention of the subject to focus on only the melodic element since he already kenw, or anticipated the rhythmic pattern which was to be used.

The major problem encountered in making the test tape for this test was the selection of an appropriate tempo at which to perform the items. The test items were presented in order of difficulty and not grouped according to the number of tones in the item. A four-tone item in $\frac{4}{4}$ following a three-tone item in $\frac{3}{4}$ might create a problem because the child might still be thinking in terms of three pulsations per measure. To minimize this difficulty it was finally decided to write all items with the quarter note as the basic unit and play the items at a tempo of ♩ = 90.

The intervals of silence which followed the aural presentation of each item were then varied in accordance with the duration of the item and the expected time required for singing a response.

This Rhythm Test, which is included in Appendix C, was then given to a group of children who had already had the 45-Item Test. This second testing session took place two months later and it appears reasonable to assume that few children would be able to remember any of the test items and thus earn a higher score. The procedures for making the test tape were similar to those detailed for the other tests and the test was administered under essentially the same conditions as prevailed for the 45-Item Test.

Analyzing and Processing the Data

Processing the data for the various tests presented no significant problems. The tapes contained all the responses of the subjects and these responses were transferred to individual data sheets for each subject. If a response was questioned, the tape was reversed and the response replayed. Correct responses, those which duplicated the aural presentation, were indicated on the data sheets by a check mark. Incorrect responses, even if only a single tone was incorrect, were entered on the data sheets in tonic sol-fa syllables for the entire response, with the correct portions of the response circled for ease of scoring. Failure to make a response to a given item was indicated by a dash on the data sheet.

The scoring system varied slightly from test to test and so each test will be discussed separately.

1. 60-Item Test Scoring

The purpose of this test has already been described as purely exploratory, with the data being used in an item analysis. Scoring of the items was done on a simple "right" or wrong" basis with no credit allowed for responses which were partially correct. One point was given for each item correctly reproduced, giving 60 points as the highest possible total score.

2. Scoring the 45-Item Test

Processing the data from the 60-Item Test soon showed the need for a system of scoring which would reward partially correct responses. This was of vital importance since the study sought to identify various levels of accomplishment, an objective impossible to realize if only perfect responses were tabulated. The data of the Exploratory Test showed that children made the following types of responses to the aural presentation of an item:

a. The response was an exact imitation of the stimulus, therefore it was correct.

b. Part of the response was an exact imitation, the remainder either being similar to the stimulus, different from the stimulus, or omitted.

c. The response was given in a different tonality from the stimulus but was correct in terms of the tonal relationships within the item. For example, C E G might be sung as D F♯ A because the subject failed to begin on the correct tone. The response was, however, still a major triad.

d. The response duplicated the contour and number of tones in the stimulus but was not an exact imitation.

e. The response showed awareness of either the contour of the stimulus or of the number of tones contained in the stimulus but any resemblance to the original presentation was accidental.
f. The response was completely unrelated to the stimulus in terms of contour, number of tones reproduced, or pitch.
g. The response, if any was made, was not given in terms of some fixed pitch, indicating that the subject was unable to use his voice for this type of task, if indeed he could sing at all, or that such a response was the result of poor memory.

Since it may be assumed that the auditory perception of musical sounds functions at several levels similar to those identified by this analysis of pupil responses, a scoring system was developed which gave credit for any response which appeared to indicate that the subject was aware of the nature of the stimulus. The point values assigned to these various scoring possibilities was arbitrarily determined by the writer since few such recommendations appear in the professional literature surveyed. This scoring system is explained as follows:

a. The response is an exact imitation of the stimulus-2 points for each tone of the item. A correct 7-tone pattern consequently would be scored 14 points.
b. Part of the response is correct, the remainder is changed or omitted - 2 points for each correct tone, but scored in terms of the sequence of tones of the stimulus. For example, the stimulus "do re mi fa so" might result in a response of "do re mi so la".

Only the first three tones are correct, the other two being changed, and so the item is scored 6 points. Another type of response to this same stimulus might be "re fa mi so do", which contains all of the tones of the stimulus but they are not in the correct sequence. In this case the response would be given only 2 points for the third tone, "mi", which appears in the proper place. A third type of response to this stimulus might be merely "mi fa so". Such a response could be given no points in terms of correct tones wven though the the final three tones of the stimulus are correctly reproduced. In such instances there appeared to be no way of determining whether the subject forgot the beginning of the item and retained the final portion, whether he retained the beginning and started on the wrong pitch but forgot the remainder of the item, or whether he thought the item contained only three tones. To eliminate the need for guessing as to the cause of such a response, the writer scored all such responses in terms of the sequence of correct tones, a decision which undoubtedly penalized some subjects but which also rewarded others.

c. The interval relationships between the tones of the response are correct, but the entire response is sung in a different tonality - 50% of the total possible points for the item. The subject hearing "do re mi" might lose the starting pitch and sing "re mi fi", an exact transposition of the stimulus, and the response would be scored 3 points, 5 % of the total item value of 6 points. An exact transposition of the stimulus is required if the item was to be scored in this way.

d. The response duplicates the general direction or shape of the item and has the correct number of tones - 25% of the total possible value of the item. The stimulus "do mi do" might result in a value of the item. The stimulus "do mi do" might result in a response of "mi fa mi" having the correct number of tones and a shape similar to the stimulus and would be given 1.5 points, or 25% of the possible total of 6 points.

e. The response contains the same number of tones as the stimulus but the pitches and/or direction are different - 12.5% of the total possible points for the item. For example, a response of "mi re do" to a stimulus of "do mi so" would be given .75 points.

f. The response follows the general direction of the stimulus but the tones are incorrect both in terms of pitch and number - 12.5% of the total possible points for the item. For example, a response of "do mi so mi" to the stimulus "la ti do la so" appears to show a recognition of the direction of the item and is scored as 1.25 points, or 12.5% of the total 1- points for the 5-tone item.

The writer was unable to discover any significant weaknesses in this scoring system although certain responses indicated that they might be scored in one of two ways. For example, Item 16, "do mi so", was frequently sung as "do re mi". The first tone of the response was correct and the response could properly be scored 2 points on this basis. However, the response could also be considered as the type described in "d" above and could be scored as only 1.5 points. The differences between these two scores is relatively small and would not materially affect the subject's position relative to others in his group in view of the fact that the highest possible total score was 460 points. In such questionable cases the writer was forced to decide the true nature of the response and, in the case of the example cited, he considered it more appropriate to score the response as 1.5 points, feeling that this was a "d" response rather than a "b" response, with the subject being more aware of the shape and number of tones presented than of the exact pitches of the stimulus. Such judgments also made it possible to more accurately identify responses by type for later analysis.

The total test score was obtained by summing the scores for each item and reducing the final figure to three digits by rounding off the decimals.

3. Scoring the 20-Item Test

All items in this test were scored on the same basis as the 45-Item Test. The total score for each trial of the test was obtained by summing the items, rounding of the decimal so the score was expressed in three digits. The mean test score was obtained by summing the scores for each trial and dividing by the number of trials, three. The highest possible score for any trial, and consequently for the total test, was 200 points.

The primary function of this test was to obtain data regarding the consistency of the subject's responses to the stimuli. The average score of the three trials would indicate such consistency provided one is willing to accept the assumption that high scores imply greater consistency than low scores. A weakness of such an assumption becomes apparent when one considers that the score regards only correct responses and does not take into account the subject who may have consistent responses which are incorrect. For example, Subject A might sing only five of the twenty items correctly but he repeats this for each of the three trials. His total test score might be relatively low but for these five items he was consistent 100% of the time. It is likewise possible that Subject B might sing the same response to a given item in each of the three trials but the response would not be correct. His score would be low but he would also be consistent 100% of the time for such items.

The writer was unable to develop a scoring formula which would be capable of indicating consistency of response accurately and so a detailed analysis of all responses was carried forward in order to give such information.

4. Scoring the Phrase Test

The nature of this test precluded the possibility of using the same scoring system as that developed for the other tests. The purpose of the test was to obtain data regarding the rate at which children were able to learn musical material when the basis for such learning depended upon the accuracy of aural perception. A second purpose was to obtain data regarding the accuracy of responses which also reflected the accuracy of aural perception. The subject was able to hear the phrase several times and, to reduce the complexity of scoring and eliminate the need for making subjective judgments, it was decided to score only on the basis of correct tones reproduced in each measure rather than to attempt to reward all partially correct responses in some fashion. There were a total of 15 tones in each of the phrases, each correct tone being scored 2 points, making 30 points the highest possible score for any trial.

The data for the test was processed in the same way as for the other tests, the pupil responses being transferred from the tape to data sheets by entering the tonic sol-fa syllables for these responses or by using a check mark if the tones in an entire measure were correct. Two kinds of scores were calculated for the Phrase Test:

a. The "mean correct" score for the entire phrase, with a possible score range of zero to 30. This score represented the achievement of the subject in terms of the number of tones which were correctly reproduced. The formula followed was: "mean correct" $= \frac{\text{sum of the scores for each trial}}{\text{number of trials}}$.

b. A "rate-of-learning" score (RL) for the entire phrase represented a single score which could describe the test gains made by the subject from the first trial to the last, as well as a score which would take into consideration the number of trials necessary to make those gains. The test gain was obtained by using the following formula:

$$\text{test gain} = \frac{a - b}{c - b}$$

a = the score on the final trial, this final trial identified as either the first of two consecutively perfect trials, or the last trial of the test if the subject was unable to score a perfect trial prior to that time.

b= the score on the first trial, which was the first singing trial and followed the second aural presentation of the phrase.

c= the perfect score, 30 points assigned to any trial for which the response was entirely correct.

The actual gain on the test which was made by the subject is expressed in the numerator, "a - b"; the total possible gain for the subject is expressed in the denominator, " c - b". The result, which can never exceed 1, is the test gain and is actually a proportion. It will be noted that the basis for determining the gain is the score on the first singing trial taken by the subject. Since the subject had had an opportunity to hear the phrase twice before attempting to sing a response, it was decided to use the score of the first trial rather than to assume a starting point of zero.

The test gain does not reflect the number of learning trials taken by the subject and, in order to obtain such a score, the test gain was divided by the number of learning trials which the subject had and the result then multiplied by 100 so the score could be expressed in larger units. The complete formula for this RL score was: $RL = \frac{a - b}{c - b} \times \frac{100}{N}$

The RL scores ranged from zero to 100 but did not include any negative scores. The writer would like to identify certain typical scoring situations in order to clarify the scoring procedure, as well as to point out a weakness which was encountered in the process.

a. The subject whose final trial represented no gain over the score on the initial trial obviously made no gain and, in effect, had a loss but these individuals were also given an RL score of zero to eliminate the problems inherent in working with negative values.

b. The criterion of success consisted of two consecutively perfect performances and, since the subject had had the opportunity to hear the phrase twice before attempting to sing, it would be possible for a subject's first and second singing trials to be perfect. Since such subjects would not be able to improve their performance, application of the scoring formula would result in an RL score of zero. Such a score seemed illogical in view of the superiority of subjects as compared to the subjects mentioned in "a" above and, to differentiate between such cases, it was decided to give an RL score of 100 to those subjects having perfect first and second trials.

c. The RL score, since it was a gain score rather than an achievement score worked to the disadvantage of subjects who may have had high scores on the initial trial, in terms of the number of correct tones, but then failed to impoove or required several trials to attain a perfect performance. For example, Subject A had a score of 28 on the initial trial and a score of 28 on the final trial, the ninth, giving an RL score of zero ($\frac{28-28}{30-28} \times \frac{100}{9} = \frac{0}{2} \times \frac{100}{9} = 0$), Subject B, on the other hand had a score of 4 on both the initial and final trials and his RL score was also a zero since no gain had occurred. These two subjects represented completely different levels of accomplishment in terms of the number of tones they were able to sing correctly but shared a common inability to improve on their initial trial. The RL score did not, and could not, differentiate between two such cases and there seemed to be no practical method by which the RL score could be combined with the "mean correct" score to obtain a single descriptive score. For this reason, the results of the Phrase Test are discussed in terms of both types of scores.

5. Scoring the Rhythm Test

The purpose of the Rhythm Test was to ascertain the possible effect the added element of rhythm might have upon the auditory perception of tonal configurations. It will be remembered that the Rhythm Test was given to a group of subjects who had earlier taken the 45-Item Test and, since the tonal patterns were the same in both tests, the scores on the two tests could be easily compared if similar scoring methods were used. It was assumed that the melodic character of the tonal pattern was not significantly altered by adding the element of rhythm although the fact that the subject had to focus his attention on both elements present in the item might well serve to reduce the effectiveness of auditory perception. The Rhythm Test was scored in terms of the number of correct, or partially correct responses of the tonal patterns rather than in terms of both rhythmic accuracy and pitch accuracy, the scoring being exactly the same as outlined for the 45-Item Test and ignoring the rhythmic element entirely. The highest possible score was again 460 points and represented only the accuracy with which the subject responded to the melodic element of each item.

The methods used to process the test data were identical with those already described for the 45-Item Test. Processing the data for this Rhythm

Test gave the writer much valuable information regarding possible scoring methods which could be applied to such tests where the concern would be with determining the accuracy of response to rhythmic elements. The processing also resulted in suggesting improved research techniques which could be employed in subsequent investigations.

These scoring methods which were employed for each of the several tests have been discussed in some detail because this constitutes a major problem in carrying forward research in the area of music. It is of the greatest importance that the scores obtained for a given test reflect both the nature of the task involved and the kind of thing the test is designed to measure. Music educators have consistently defended the need for subjectively scoring the many aspects of musical accomplishment and development on the grounds that the things being measured could not be scored objectively because of the unique characteristics of music itself. The writer makes no claim to having satisfactorily solved the problem of objectively measuring the several aspects of a subjective art. He does feel, however, that continuing research activity will eventually result in a much greater degree of objectivity than has heretofor been believed possible. Such objectivity will, in turn, have far-reaching effects upon the musical instruction taking place at all levels of our educational system.

CHAPTER IV

RESULTS OF THE STUDY

Introduction

A total of 660 children were scheduled for testing during the course of the project, this number including the original sample described in Table IV, page 25, as well as alternates. These children were to be given the basic 45-Item Test and one of the three other tests which were to be used. Since it was impossible to administer all four tests to each of subjects, the 660 children were randomly assigned to one of three groups of approximately equal size. Following the assignment to groups, the Rhythm Test, Phrase Test, or 20-Item Test was randomly assigned to one of the three groups. Throughout the testing period it became quite evident that it would be impossible to secure data for all of the children. Certain children were not available for testing at the scheduled time and the test results for other children were judged not usable. In reporting the results of the study, the data for each of the several tests will be discussed separately to facilitate the presentation.

Musical Experience

The Information Sheet (Appendix B) was designed to collect certain data regarding the musical training of the children, as well as data which described the kinds and extent of musical activities which might take place in the home. The basic purpose of this data was to permit the identification of two groups of children: Group I, those with high "musical experience scores"; and Group II, those with "low musical experience scores". This "experience score" was a composite of musical training for the child and the kinds and amounts of musical activity which the child reported taking place at home. The score was obtained by summing the weighted scores for each item. The scoring key used for this purpose is included with the Information Sheet in Appendix B.

Information Sheets were returned by 506 children of the total sample and, although complete data was not available for all of the children taking each of the tests, it was possible to identify the upper and lower groups in each grade. To establish the score ranges for each group, a frequency distribution was set up and the lower and upper quarters identified; $Q_1 = 16.5$, $Q_3 = 29.5$. Children with scores between 0 and 16.5 were assigned to Group I; those with scores of 29.5 or higher were assigned to Group II; and those with scores in the middle range were not assigned to either group. Therefore, in subsequent discussion of this variable, Experience Group I is the group with below average musical experience score; Experience Group II is the above average group.

The data for each of the items on the Information Sheet is not relevant to the report and is summarized in Appendix B. Subsequent references to the experience groups will be made when discussing the results of the several tests.

45-Item Test

Usable data on this basic test was obtained for 606 children, distributed by school, grade, and sex as shown in Table V, page 28. For reasons given elsehwere, it proved impossible to maintain an equal proportion of boys and girls at each grade level, or to maintain the original distribution for each grade in each school.

One of the first steps taken with regard to this 45-Item Test was to ascertain the reliability of the test itself. A detailed item analysis for each grade of the total sample was carried forward and the procedures need not be detailed because they are generally known. The Kuder Richardson No. 20 formula was used to compute reliability coefficient of the test:

$$r_{tt} = \frac{n}{n-1} \left[\frac{\sigma^2_t - \Sigma pq}{\sigma^2_t} \right]$$

To compute the standard error of measurement, the usual formula was used:

$$\sigma_e = \sigma_t \sqrt{1-r_{tt}}$$

The data presented in Table X indicates that the test possessed a high degree of stability and internal consistency at all grade levels.

TABLE X

Reliability Coefficients and Standard Errors
for the 45-Item Test

Grade	r_{tt}	σ_e	Grade	r_{tt}	σ_e
1	.95	2.29	5	.97	2.31
2	.96	2.34	6	.97	2.34
3	.96	2.39	1-6	.97	2.36
4	.97	2.37			

It will be recalled that the scoring system used for the 45-Item Test involved weighted scores, the total possible value of the item being determined by the length of that item. This scoring system also gave recognition to responses which were partially correct. However, in order to simplify calculating procedures and still secure the desired information, these reliability coefficients were derived from data using unweighted scores, one point for each item correct, and did not take into account the partially correct responses. This gave a score range of 0 to 45 which coincided with the number of test items.

One purpose of the test was to determine whether any significant differences were found between grade levels. The data in Table XI shows a higher mean score at each successive grade level, with a marked increase for the 6th grade. To ascertain the significance of these observed differences, "t" tests were made between grades with the results summarized as follows:

Grades	"t"	Grades	"t"	Grades	"t"
1 and 2	1.583	1 and 3	2.065*	1 and 4	2.861**
2 and 3	0.601	2 and 4	1.318	2 and 5	1.971*
3 and 4	0.634	3 and 5	4.811**	3 and 6	5.281**
4 and 5	4.114**				

* Significant at the 5% level.
** Significant at the 1% level.

There are significant differences between grades 5 and 6 when we consider the grades at one-year intervals. When a two-year interval is taken we find significant differences between grades 1 and 3 and between

TABLE XI

Summary, by Grade, of Scores on 45-Item Test

GRADE	N	EX	EX^2	$\bar{X}$	σ
1	120	23,121	5,850,861	192.7	107.9
2	103	22,202	5,947,182	215.6	106.2
3	93	20,962	6,105,108	225.4	121.8
4	102	24,109	7,130,739	236.4	118.5
5	102	25,199	7,719,369	247.0	121.0
6	86	27,190	9,551,982	316.2	105.4
1-6	606	142,763	42,305,241	235.6	119.6

TABLE XII

Frequency Distribution, by Grade, of Scores on the 45-Item Test with Reference to Quartiles for the Total Sample

GRADE	(0 - 140.5) N	%	Q_1 (141 - 226) N	%	Q_2 (226.5 - 332.0) N	%	Q_3 (332.5 - 460) N	%
1	43	36	34	28	28	23	15	12
2	30	29	28	27	29	28	16	16
3	23	25	26	28	24	26	20	22
4	26	26	24	24	24	24	28	28
5	23	23	26	26	24	24	29	28
6	6	7	15	16	24	28	41	48

Q_1 Q_2 Q_3

(141) (226.5) (332.5)

4 and 6. A three-year interval shows significant differences for all grades compared. It would appear that while children in the middle grades, 3, 4, and 5, perform somewhat alike, there are marked differences between the lower and the upper grades.

The relatively large standard deviations for each grade indicate that individual scores are distributed over a fairly wide range. Inspection of the raw data showed the following score ranges for each grade:

Grade 1, 12 to 401	Grade 4, 26 to 450
Grade 2, 31 to 458	Grade 5, 20 to 460
Grade 3, 9 to 460	Grade 6, 82 to 460

Another basis for comparing the performances of the children in each of the grade levels makes use of the frequency distribution of the scores. The six grades were combined into a single distribution for this purpose and the 25th, median, and 75th percentiles were located for this total sample. This gave $Q_1 = 141$; $Q_2 = 226.5$, and $Q_3 = 332.5$. When the score values had been established for the total group of 606 children, it became possible to compare the performances at each grade level, identifying competence levels for the test. Table XII summarizes this information, showing the number and percentage of children earning scores located below Q_1; between Q_1 and Q_2; between Q_2 and Q_3; and above Q_3. The percentage of scores which lie within the inter-quartile range tends to remain constant from grade level to grade level, but there is a pronounced increase in the proportion of children earning scores above Q_3 as we examine each successively higher grade level.

Table XII is to be read as follows: for 1st grade 43, or 36% of the 120 children had scores between 0 and 141, the lower quartele of the total distribution; 15, or 12% had scores between 332.5 and 460, above the upper quartele of the total distribution. All other entries are to be interpreted in like manner, keeping in mind that Q_1, Q_2 and Q_3 were established for the total sample of 606, not for each of the several grade levels.

This would further tend to support the position that age (grade level) implies a musical growth and maturity which is directly related to competence in the auditory perception of the melodic elements of music.

One of the instructional problems which faces the teacher of music results from the markedly different attitudes which boys and girls have toward music. There is no sound basis for assuming that these differences can be attributed to differing abilities. In a previous study (10) the writer found no significant differences between boys and girls as they engaged in certain music reading activities. To determine whether such differences might exist with the auditory perception of musical sounds, the writer compared the performances of boys and girls at each grade level. The data in Table XIII shows the girls to have a higher mean than the boys at all grade levels. The "t" tests between the means of the boys and girls show the following.

Grade	"t"	Grade	"t"
1,BG	0.664	5,BG	0.508
2,BG	1.565	6,BG	0.022
3,BG	1.189	1-6,BG	1.190
4,BG	0.281		

None of these were significant at the 5% level and the hypothesis "no difference" may not be rejected, results which confirm the writer's previous

TABLE XIII

Summary, by Sex, of Scores on the 45-Item Test

Grade	N		ΣX		ΣX^2		$\overline{X}$		σ	
	B	G	B	G	B	G	B	G	B	G
1	62	58	11,551	11,570	2,863,435	2,987,426	186.3	199.5	107.1	108.2
2	43	60	8,439	13,763	2,073,537	3,873,645	196.3	229.4	98.5	109.3
3	51	42	10,798	10,164	3,064,772	3,040,336	211.7	242.0	123.6	117.6
4	47	55	10,940	13,169	3,244,582	3,886,157	232.8	239.4	121.9	115.4
5	42	60	10,068	15,131	2,980,982	4,738,387	239.7	252.2	116.2	124.0
6	52	34	16,430	10,760	5,753,054	3,798,928	316.0	316.5	103.9	107.6
1-6	297	309	68,226	74,557	19,980.362	22,324.879	229.7	241.3	120.4	118.5

findings.

Programs of musical activities are frequently influenced by an assumption that additional musical training and out-of school contact with musical activities in the home is significantly related to the musical growth and development of the individual. Since the tests used in this study required a response in terms of a musical performance it is reasonable to expect that these varying kinds and amounts of musical experiences might be reflected in the children's test scores. The data summarized in Table XIV clearly shows that such is the case. To further test the hypothesis of no difference between experience groups, "t" tests were made between the two groups at each grade level, with the results summarized as follows:

Grade	"t"	Grade	"t"
1, I II	3.222**	5, I II	3.642**
2, I II	2.512**	6, I II	2.334*
3, I II	3.290**	1-6, I II	7.909**

* Significant at the 5% level.
** Significant at the 1% level.

All differences are significant at the 1% level except for grades 2 and 6, which are significant at the 5% level. It would appear that the hypothesis of "no difference" could be rejected, although other hypotheses regarding the nature of this relationship between musical experience and musical achievement, in terms of auditory perception, should be tested.

The point of view is generally accepted that the kinds and amounts of out-of-school musical experiences are related to the socio-economic back-

TABLE XIV

Summary, by Experience Groups, of Scores on the 45-Item Test

Grade	N		ΣX		ΣX^2		$\bar{X}$		σ	
	I	II	I	II	I	II	I	II	I	II
1	27	22	4,276	5,563	940,636	1,633,235	158.4	252.9	98.8	101.5
2	27	23	4,660	5,563	982,158	1,621.241	172.6	241.9	81.2	109.5
3	29	23	5,454	6,813	1,456,620	2,280,091	188.1	296.2	121.9	106.7
4	32	28	6,490	8,086	1,603,032	2,721,810	202.8	288.8	94.7	117.5
5	22	38	4,218	11,524	1,095,622	3,965,862	191.7	303.3	114.2	111.3
6	14	24	3,810	8,579	1,156,662	3,372,137	272.1	357.5	92.5	112.8
1-6	151	158	28,908	46,128	7,234,730	15,594,376	191.4	291.9	106.1	116.0

Experience Groups: I = Low Experience Scores; II = High Experience Scores.

grounds of children. Empirical information repeatedly suggests that musical activities are more likely to be recognized as important in homes where the occupational status of the parent is professional or semi-professional, and where there is a relatively high economic level. Such an environment is certain to be reflected in the child's own choice of activities and presumably also his familiarity with music. The present study did not attempt to explore this relationship in detail because of the complexity of gathering relevant data and the difficulty of accurately defining the many possible variables. The writer was, however, interested in ascertaining whether the hypotheses of "no difference" between children from high socio-economic backgrounds was tenable. The writer consulted with authorities of the Madison school system and it appeared possible to assign certain schools to one of these two socio-economic groups. Other schools, because of the nature of the pupil population, could not so clearly be categorized. The writer is aware of the obvious limitations of such procedures, but the results suggest that it might be well to further explore the problem. The two groups identified by this estimate were: Group I, school populations representing a high socio-economic level; and Group II, school populations representing a low socio-economic level. No attempt was made to accurately identify the exact limits to be observed in regard to family income, education of parents, occupation of parents, and other similar factors which should be considered in further testing hypotheses of this kind.

The six grades were combined for this purpose and the means of Group I and II were compared. For the 45-Item Test, the mean of Group I was 261.3; for Group II, 222.4. The σ of Group I was 120.1; for Group II, 117.6. The "t" value of 3.361 was significant at the 5% level and the hypothesis of "no difference" can be rejected with certain reservations. Further checking at each of the several grade levels, using more precise data-gathering devices is necessary if we are to adequately test this, and other related hypotheses. The data does appear to indicate that the environment of children is related to their competence in music, at least as regards tasks of the kind used in the present study.

The reader will recall that a very detailed scoring system was used which recognized several types of partially correct responses. The raw data was subjected to an extensive item analysis, partially to provide data for the reliability coefficients and to ascertain the characteristics of pupil responses at each grade level. The data obtained is too extensive to be included in a report of this kind and a discussion of each test item does not appear to be particularily relevant. The general characteristics may be summarized by combining all items of the test and discussing the total test in terms of the proportions of children making certain types of responses. The general types were as follows:

A. The item is correct, the response is an exact imitation of the stimulus.

B. Part of the response is an exact imitation of the stimulus.

C. The item has been transposed to another key, and although the tonal relationships are correct, the pitches are incorrect as compared to the stimulus.

D. The response duplicates the shape of the stimulus and contains the correct number of tones.

E. The response either (1) follows the general direction of the stimulus but does not contain the correct number of tones or the correct pitches; or (2) the response contains only the correct number of tones of the stimulus but the pitches and direction are incorrect.

F. Totally incorrect response or no attempt made.

Table XV summarizes the percentage of children in each grade making each of these six general types of responses. Columns A through F identify the types of responses just described.

TABLE XV

Summary of Types of Responses made by children to the 45-Item Test

Grade	A	B	C	D	E	F
1	24%	26%	4%	12%	15%	19%
2	29	25	5	12	16	13
3	33	24	4	11	13	15
4	35	22	5	13	13	12
5	38	22	5	11	13	11
6	55	21	4	8	6	5
1-6	35	24	4	12	13	13

The data shows quite clearly that response types A and B account for more than half of the kinds of responses at each grade level, indicating that half of the children are capable of correct or partially correct responses to aural presentation of an item. Response C shows that few children are able to accurately transpose the stimulus to a different tonality. Responses D and E remain somewhat constant from grade to grade, except for 6th grade, and show that these children are aware of the shape and content of the stimulus even though they are unable to control the voice sufficiently to reproduce the stimulus correctly. This group, accounting for approximately 25 per cent of the children in the first five grades, needs to be given considerably more help and attention than they are presently receiving before they will be able to use the singing voice effectively. The percentage of children who were either unable to retain the stimulus long enough to attempt a response or whose response was completely unlike the stimulus also tends to remain somewhat constant from grade 2 through grade 5. In fact, the summary clearly shows that for any one of these six types of responses the proportion of children making that response tends to remain almost constant throughout the first five grades. The writer is unwilling to accept the hypothesis that a sudden improvement takes place between the fifth and sixth grade but rather that these data indicate the result of randomly selecting a large number of sixth grade children who were musically superior. This analysis further suggests that many children, particularily those making Type D, E and F responses, do not change from grade 1 through grade 5. The writer's projected longitudinal study is expected to give data which will permit the testing of such an hypothesis.

Summary of the Findings

1. The reliability coefficients for the test were uniformly high at each grade level, indicating a high degree of stability and internal consistency.
2. The "t" tests do not justify rejecting the hypothesis of "no difference"

between the means of two grades, when these grades are taken at one-year intervals. When a two-year interval is taken, the hypothesis may be rejected for the grades 1 and 3 and grades 4 and 6. A three-year interval, which clearly separates the lower three grades from the upper three grades shows that the differences are highly significant, that these two levels apparently differ in terms of this kind of task.

3. The competence levels established in terms of the total sample clearly indicate that there are individuals within each grade whose performance is above Q_3, and that this proportion tends to increase at each successive grade level. The data cannot indicate the reason for such shifting, nor can it describe the amount of improvement which takes place as the child becomes older and more musically experienced. Except for the sixth grade, the proportion of children at each of the other three competence levels tends to remain constant from grade level to grade level.
4. None of the "t" tests showed that boys and girls differed significantly in their ability to perceive and respond to the auditory presentation of melodic items.
5. All of the "t" tests showed that above-average musical training and out-of-school musical experiences resulted in means which were significantly higher than the means of children having considerably more limited experience.
6. The relationship between socio-economic background and musical competence appears to be significant when all grades are combined since the difference between the two means was significant at the 5% level.
7. The analysis of the types of responses made by children showed that more

than half of the children at each grade level were able to make correct or partially correct responses to the aural presentation of an item. Approximately 14 per cent of the children in each of the first five grades were unable to give a response which indicated any degree of awareness of the nature of the auditory stimulus. This analysis also showed that the proportion of children making each type of response tends to remain relatively constant throughout the first five grades.

Rhythm Test

This test was, it may be recalled, similar to the 45-Item Test in that it used the same test items with these items presented rhythmically as well as melodically. Approximately 200 children had been scheduled to receive the Rhythm Test. However, since many children were not available for testing, usable data for the test was obtained for 165 children. This data is summarized in Table XVI and, as was the case with the 45-Item Test, there are higher mean scores for each successive grade level, except at grades 2 and 3. The "t" tests between grades are:

Grades	"t"	Grades	"t"	Grades	"t"
1 and 2	1.821	1 and 3	1.445	1 and 4	2.532**
2 and 3	0.267	2 and 4	0.642	2 and 5	1.084
3 and 4	0.895	3 and 5	1.280	3 and 6	3.754***
4 and 5	0.469	4 and 6	2.927**		
5 and 6	2.153*				

* Significant at the 5% level.
** Significant at the 2% level.
*** Significant at the 1% level.

TABLE XVI

Summary by Grade of Scores for Children Taking the Rhythm and 45-Item Tests

Grade	N		ΣX		ΣX^2		$\bar{X}$		σ	
	1	2	1	2	1	2	1	2	1	2
1	35	35	6,608	6,085	1,641,524	1,426,921	188.8	173.9	106.1	102.7
2	30	30	7,127	6,521	2,159,961	1,831,975	237.6	217.4	124.7	117.5
3	24	24	5,482	4,911	1,599,834	1,299,979	228.4	204.6	120.4	110.9
4	22	22	5,699	5,255	1,720,045	1,519,175	259.0	238.9	105.3	109.5
5	26	26	7,179	6,823	2,466,103	2,195,261	276.1	262.4	136.4	124.8
6	27	27	9,435	9 152	9,596,609	3,373,424	349.4	339.0	105.3	100.2
1-6	164	164	41,530	38,747	13,184,076	11,646,735	253.2	236.3	127.5	123.3

Rhythm Test = 1; 45-Item Test = 2.

These show that the differences are significant in only four of the twelve comparisons, with three of these involving the 6th grade. It appears that the hypotheses of "no difference" is tenable for all grades but this exceptional 6th grade and grades 1 and 4. Had the 5th grade performed better on this test the hypotheses of no difference might have been rejected when the lower grades were compared with the upper grades.

This parallelism with the "t" tests made for the 45-Item Test is to be anticipated since these two tests had essentially the same content. The reader may also remember that the scoring system was the same for both tests, the items scored in terms of melodic accuracy, with rhythmic accuracy not being considered. The purpose for scoring the Rhythm Test in this manner was to permit comparisons between the two tests in an attempt to ascertain whether the introduction of the rhythmic element significantly affected the auditory perception of the melodic element. Table XVI gives the data for both the Rhythm Test and the 45-Item Test for this group of children.

Correlations between the two test scores showed that a high relationship exists. The correlations were as follows:

Grade	"r"	Grade	"r"
1	.92	4	.94
2	.94	5	.96
3	.96	6	.94

This indicates that the two tasks are very closely related, as would be expected, and that children could be expected to perform at much the same level of competence on either test.

The data in Table XVI shows that the mean scores for the 45-Item Test are lower than those for the Rhythm Test at every grade level as well as for all grades combined. Since the Rhythm Test aas given at least two months later than the 45-Item Test it is not reasonable to assume that children were able to remember the test items. This higher means for the Rhythm Test might reflect the influence of previous practice and familiarity with the task; or they might suggest that the rhythmic organization of the melodic items makes them easier to perceive and retain. Tests for the significance of the differences between the means of the two tests (I = Rhythm, II = 45-Item) showed the following:

Grade	"t"	Grade	"t"
1, I II	0.588	5, I II	0.370
2, I II	0.635	6, I II	0.365
3, I II	0.698	1-6, I II	1.215
4, I II	0.603		

None of the "t" values are significant at the 5% level and the hypothesis of no difference between the means of the two tests is tenable. This would indicate that the addition of the element of rhythm to melodic items did not have a significant influence upon the auditory perception of such items; that such an addition proved to neither a help nor a handicap. Therefore the higher means for the Rhythm Test might well have resulted from the previous practice on the 45-Item Test but, lacking a control group who had taken only the Rhythm Test, there seems to be no way to test the hypothesis at this time.

The fact that this hypothesis is tenable is highly significant because it has been generally assumed that auditory perception is influenced when more than one musical element is present in the stimulus. The writer is even more convinced that the problem needs further study, that other hypotheses need to be tested.

Analysis of the data for the 45-Item Test showed that no significant differences were observed between the mean scores of boys and girls at any grade level. Although the data presented in Table XVII show again that the girls have higher mean scores than the boys, the "t" tests showed that none of these differences were significant at the 5% level.

Grade	"t"	Grade	"t"
1, BG	1.091	5, BG	1.079
2, BG	0.868	6, BG	1.320
3, BG	1.481	1-6, BG	1.228

These results were to be anticipated, although it is apparent that other factors may be present in a situation which so consistently gives higher scores to the girls. Such factors would undoubtedly include attitudes toward the task and toward music in general, level of aspiration, general singing competence, and overall motivation. The report does not, however, propose to consider these factors.

Comparisons between experience groups for children taking the Rhythm Test do not produce results similar to those found for the 45-Item Test. Part of this is undoubtedly due to the small number of children who took the Rhythm Test data was also available regarding their musical background and experience. Table XVIII shows the mean scores for all grades to be higher for Group II than for Group I, while the "t" tests showed the following:

TABLE XVII

Summary by Sex, of Scores on the Rhythm Test

Grade	N		ΣX		ΣX^2		$\bar{X}$		σ	
	B	G	B	G	B	G	B	G	B	G
1	12	24	1,882	4,769	380,016	1,263,357	156.8	198.7	84.1	114.7
2	12	18	2,554	4,573	762,778	1,397,183	212.8	254.1	135.2	114.4
3	12	12	2,306	3,176	623,692	976,142	192.2	264.7	122.7	106.3
4	8	14	2,122	3,577	632,446	1,087,599	265.3	255.5	93.3	111.4
5	12	14	2,933	4,246	957,915	1,508,188	244.4	303.3	141.7	125.5
6	17	10	5,590	3,845	2,090,016	1,506,593	328.8	384.5	121.7	53.1
1-6	73	92	17,387	24,186	5,446,863	7,739,062	238.2	262.9	133.7	122.5

TABLE XVIII

Summary, by Experience Groups, of Scores on the Rhythm Test

Grade	N		ΣX		ΣX^2		$\overline{X}$		σ	
	I	II	I	II	I	II	I	II	I	II
1	7	3	879	642	135,975	149,822	125.6	214	60.5	64.4
2	6	6	1420	1,456	382,986	441,214	236.7	242.7	88.4	121.0
3	5	6	811	1,728	136,305	576,046	162.2	288.0	30.9	114.3
4	10	8	2301	2,345	597,797	827,841	230.1	293.1	82.7	132.5
5	7	9	1330	3,312	408,442	1,291,392	190.0	368.0	149.2	89.8
6	5	10	1601	3,697	595,417	1,493,223	320.2	369.7	128.7	112.4
1-6	40	42	8342	13,180	2,256,922	4,779,538	208.6	313.8	113.7	123.8

Experience Groups: I = Low Experience Score; II = High Experience Scores.

Grade	"t"	Grade	"t"
1, I II	1.859	5, I II	2.766*
2, I II	0.090	6, I II	0.712
3, I II	2.161	1-6, I II	3.955**
4, I II	1.163		

* Significant at the 2% level.
** Sifnificant at the .01% level.

The only significant difference is observed for the 5th grade, although there is a highly significant difference for the total sample when these two groups are compared. It is highly probable that had data regarding this musical experience been available for all 165 children taking the Rhythm Test, the results of these "t" tests would have been very similar to those of the 45-Item Test. There is no basis for assuming that such experience could be related to one task and not to a second, and very similar task.

Summary of the Findings

1. The "t" tests do not justify rejecting the hypothesis of "no difference" between the means of two grades except when the 6th grade is involved.
2. The high correlations between the 45-Item and Rhythm Tests clearly indicate that children perform at approximately the same level of competence for each task. They also indicate the similarity between the two tests in terms of content.
3. The hypothesis that the addition of the element of rhythm to a melodic pattern has a significant influence upon the auditory perception of melodic patterns may, for this study, be rejected since none of the "t"

values approached the 5% level of significance. Other hypotheses concerning the relationships between these two elements of music need to be tested, other types of tests need to be constructed.

4. The "t" tests continue to show that the hypotheses of "no difference" between the means of boys and girls is tenable.
5. Musical experience does not appear to have a significant relationship to competence in performing a task of this kind. The "t" values are significant when these two experiences are compared at the fifth grade level and when all grades are combined. However, these results might well have been different had data been available for larger numbers of the sample.
6. The results of this exploratory study of rhythm gave data which suggest research procedures and techniques to be used in subsequent investigations of the auditory perception of rhythmic elements.

Phrase Test

The basic purpose of this test has been described earlier in the report and will not be repeated here. Briefly, the test was designed to provide data regarding the auditory perception of a longer, more complete melodic entity as found in the musical phrase. The task of perceiving and responding to short tonal patterns which are presented only once is somewhat different from perceiving and responding to a musical phrase consisting of 15 tones and extending through four measures. In this instance, it is unreasonable to expect children to be capable of making a successful response after a single presentation of such a phrase. Therefore, there were a total of ten presentations of the phrase, with a response to be given after the second, and each succeeding presentation. These several learning trials made it possible to identify, by means of the scoring procedures outlined in Chapter III, the rate at which children were able to learn the material, as well as the overall achievement in terms of the number of tones of the phrase which were correctly reproduced.

The reader may recall that the basis for scoring the Phrase Test was quite different than scoring the other tests. The phrase contained 15 tones and, by scoring two points for each correct tone (taken in sequence), the perfect score for any response trial was 30. The "mean correct" score was obtained by summing the scores for each trial and dividing this figure by the number of trials taken. It was impossible for any individual to have a "mean correct" score higher than 30. The RL (rate-of-learning) score described the gains made from the initial to the final trial, divided by the number f trials and, although the range was 0 to 100, a score of 11.1 or higher was possible only if the subject's final learning trial of the phrase was perfect. Consequently, the means and σ's are relatively small values, thus reflecting the scoring system which was used and the difficulty of the task. It is important to keep this in mind as the results of the Phrase Test are discussed.

Usable data was obtained for 212 children taking the Phrase Test and this is summarized in Table XIX. The data shows that once again there is improvement from grade level to grade level and, in terms of the RL scores, the 6th grade me n reflects a large number of children having perfect final trials which gave them scores of 11 and higher. There is a pronounced gap between

TABLE XIX

Summary, by Grade of Scores on the Phrase Test

A. RL Scores

GRADE	N	EX	EX^2	$\bar{X}$	σ
1	38	86.7	398.71	2.28	2.3
2	32	110.8	834.22	3.46	3.75
3	33	247.3	11,417.43	7.49	17.02
4	41	291.3	13,915.31	7.10	17.00
5	34	286.2	6,168.74	8.42	10.52
6	34	501.0	20,671.88	14.74	19.77
1-6	212	152.33	53,406.29	7.19	14.15

B. Mean Correct Scores

GRADE	N	EX	EX^2	$\bar{X}$	σ
1	38	355.9	4,541.35	9.37	5.64
2	32	303.2	3,800.94	9.48	5.39
3	33	456.7	7,557.07	13.84	6.12
4	41	453.4	6,713.10	11.06	6.44
5	34	488.2	8,910.54	14.36	7.48
6	34	569.5	11,271.25	16.75	7.14
1-6	212	262.69	42,794.25	12.39	6.95

grades 2 and 3 and between grades 5 and 6 when RL scores are considered. The same gap is observed between grades 2 and 3 when the mean correct scores are compared. The "t" tests between grades for the Phrase Test did not generally give results which were similar to those observed for the other tests.

Grades	"t" (RL)	"t" (Mean Correct)	Grades	"t" (RL)	"t" (Mean Correct)
1 and 2	1.590	0.081	2 and 4	1.173	1.104
2 and 3	1.289	3.001***	3 and 5	0.264	0.306
3 and 4	0.096	1.862	4 and 6	1.773	3.579***
4 and 5	0.387	2.026*			
5 and 6	1.621	1.329***	1 and 4	1.712	1.223
			2 and 5	2.480**	2.983***
1 and 3	1.841	3.159***	3 and 6	1.580	1.762

* Significant at the 5% level.
** Significant at the 2% level.
*** Significant at the 1% level.

These "t" tests show that for the RL scores the hypothesis of "no difference" between means is untenable for only grades 2 and 5; that for the "mean correct" scores this hypothesis is untenable for grades 2 and 3, 4 and 5, 1 and 3, 4 and 6, and 2 and 5. In other words, if the rate at which children are able to learn materials is the basis for comparison, there is, with one exception, no significant change from grade level to grade level, but when achievement is considered, the means are significantly different. in five of the twelve comparisons for this sample. The achievement scores on the Phrase Test reflect the same kinds of competencies observed for the 45-Item and Rhythm Tests, while the RL scores measure an additional factor. The writer considers it significant that this ability to learn a musical phrase from aural presentation alone does not appear to change throughout the several grades and, although the hypothesis is tenable, it needs careful examination because it does not agree with the widely accepted viewpoints of professional music educators. Other types of tests, and other hypotheses need to be developed in order to completely explore this kind of musical learning.

The raw data for the Phrase Test shows in terms of the RL scores that a definite break occurs between grades 4 and 5. Table XX gives the frequency, in percentages, of each of the several scores at each grade level. It will be noted that none of the children in grade 1 were able to learn the phrase in ten trials; but that at each successive grade level more children demonstrated this ability. The RL scores of 0 were given when no gains were made or when the score on the final trial was lower than that on the initial trial. The RL scores above 11.0 may be interpreted in this manner: 11.1 = learning the phrase after 9 trials; 12.5 = learning the phrase after 8 trials; thus decreasing the number of learning trials as the scores increase until 100 = learning the phrase after a single trial. The RL scores of 0 to 11.0 indicate that gains were made from the initial to the final trial but that no perfect trial was attained. Table XX shows that the proportion of children earning scores in this range tends to remain constant through the first three grades and then declines at each of the upper three grade levels. There is a noticeable increase in the proportion of children earning scores above 11.1 at the 5th and 6th grade levels. It is also significant to note that the proportion of children earning scores of 0 does not noticeably de-

TABLE XX

Frequency Distribution, in Percentages, of RL Scores on the Phrase Test

	Grade						
Score	1	2	3	4	5	6	1-6
100			3.0	2.4		2.9	1.4
50				2.4	2.9	5.9	1.9
33.3					2.9	2.9	0.9
25			3.0			8.8	1.9
20				4.9	2.9	5.9	2.4
16.7		3.1			8.8	11.8	3.8
14.3					5.9	2.9	1.4
12.5		3.1	3.0		1.8	2.9	2.8
11.1			3.0	4.9	5.9	5.9	3.3
.1-11.0	65.8	71.9	72.8	58.5	44.1	35.3	58.0
0	34.2	21.9	15.2	26.8	17.6	14.7	22.2
Total %	100.0	100.0	100.0	99.9	99.9	99.9	100.0

Per cent of children who learned the Phrase as shown by RL scores of 11.1 to 100.

Grade 1 - 0%

Grade 2 - 6%

Grade 3 - 12%

Grade 4 - 15%

Grade 5 - 38%

Grade 6 - 50%

Grades 1 to 6 20%

cline from grades 3 through 6, which suggests that many children who demonstrate limited competence, of the kind measured by this task, are not able to improve their level of performance after they reach grade 3. This, of course, needs to be studied more extensively before such an hypothesis could be accepted.

The Phrase Test included a second phrase which was in the minor mode. Only 25 of the 212 children were able to proceed to this second phrase within the time limits of the single testing session. These were children who had required no more than 6 trials to learn the first phrase, consequently there was additional time available for continued testing. These 25 children were distributed as follows: Grade 1-0; grade 2-1; grade 3-2; grade 4-2; grade 5-9; and grade 6-11. Of this total, 16 children were able to learn the second phrase, earning RL scores of 11.1 or higher.

The "mean correct" scores described the achievement of the individual in terms of the mean number of tones correctly reproduced. It was quite possible, therefore, that a subject might have a relatively high "mean correct" score but, because little improvement took place between the initial and final trials, his RL score could be very low. One way to identify this type of relationship was to obtain the correlation between the two kinds of scores: (1) RL score; (2) "mean correct" score. These correlations are as follows:

Grade 1, $r_{12} = .28$
Grade 2, $r_{12} = .78$
Grade 3, $r_{12} = .59$
Grade 4, $r_{12} = .72$
Grade 5, $r_{12} = .74$
Grade 6, $r_{12} = .64$

These show that many children, particularily at the first grade level, were able to sing part of the phrase correctly but were unable to improve on this performance throughout the several learning trials. The exact nature of this relationship cannot be determined from the data available but it is quite probable that the correlations would have been considerably higher had the testing procedures involved more than listening to the phrase. Children might have learned the phrase more rapidly and effectively had one or more of the following devices been employed:

a. The examiner describes and demonstrates the nature of the pupil's error.

b. The pupil has an opportunity to hear his response played back so he can compare it to the stimulus.

c. Drill on troublesome intervals is carried forward before going on to next trial.

d. The first half of the phrase is learned before going on to the second half.

The writer was not, however, concerned with using such devices. The purpose was to ascertain the amount of learning which would result from only the aural perception of the stimulus. The degree to which the child is able to perceive, retain, organize a response, and then compare this response to the stimulus depends upon the level at which his auditory perception functions. To introduce external devices would have permitted the child to learn from more than just his own auditory perception of the phrase.

The "mean correct" scores, because they do not directly describe the amount of learning taking place, show a much higher correlation with the scores earned on the 45-Item Test than do the RL scores. These correlations are keyed as follows 1 = 45-Item Test; 2 = RL score on Phrase Test; 3 = "mean correct" score on Phrase Test.

Grade	r_{12}	r_{13}	Grade	r_{12}	r_{13}
1	.28	.79	5	.66	.86
2	.42	.57	6	.54	.70
3	.38	.79	1-6	.47	.74

In every instance the r_{13} is higher than the r_{12}, and all but the first and third grade r_{12}'s are significant at the 1% level.

The data for the two tests which have previously been discussed showed that the differences between the means of the boys and girls were not significant at any grade level. The same findings were noted for the Phrase Test and only at the fourth grade level was the mean for the girls not higher than that for the boys.

The data is summarized in Table XXI. The "t" tests for the significance of the difference between the means of the boys and girls were as follows:

Grade	"t" (RL)	"t" (Mean Correct)	Grade	"t" (RL)	"t" (Mean Correct)
1, BG	1.428	1.256	5, BG	0.557	0.789
2, BG	0.434	0.546	6, BG	0.681	0.031
3, BG	1.051	0.620	1-6, BG	0.426	0.544
4, BG	1.528	1.288			

None of these met, or even approached, the criterion of significance at the 5% level and the hypothesis of "no difference" continues to be tenable.

Table XXII summarizes the scores for the Phrase Test in terms of the experience groups. Unfortunately, data regarding this experience was ob-

TABLE XXI

Summary, by Sex, of Scores on the Phrase Test

Grade	N		$\overline{X}$ (RL Score)		σ (RL Score)		$\overline{X}$ (Mean Correct)		σ (Mean Correct)	
	B	G	B	G	B	G	B	G	B	G
1	24	14	1.88	2.98	1.92	2.70	8.48	10.88	4.97	6.87
2	15	17	3.15	3.74	2.77	4.42	8.91	9.98	5.25	5.46
3	19	14	4.78	11.17	5.78	24.79	13.26	14.63	5.82	6.43
4	19	22	11.45	3.35	23.75	4.57	12.46	9.85	7.53	5.01
5	13	21	7.11	9.23	13.16	8.38	13.05	15.17	6.91	7.70
6	21	13	16.6	11.73	22.8	12.94	16.78	16.70	7.39	6.71
1-6	111	101	7.58	6.75	15.87	11.97	12.14	12.66	7.01	6.87

TABLE XXII

Summary, by Experience Groups, of Scores on the Phrase Test

Grade	N		$\overline{X}$ (RL Score)		σ (RL Score)		$\overline{X}$ (Mean Correct)		σ (Mean Correct)	
	I	II	I	II	I	II	I	II	I	II
1	9	9	3.0	1.96	2.27	1.43	8.90	11.86	5.23	4.25
2	9	9	1.79	6.01	2.12	5.26	6.13	13.77	2.61	6.72
3	13	9	4.17	14.60	3.15	30.3	11.85	16.57	5.93	6.46
4	11	15	3.56	13.41	2.93	26.38	10.95	12.63	4.74	7.57
5	5	14	3.40	11.24	4.68	12.25	10.74	17.18	7.13	6.08
6	5	11	13.90	21.69	19.0	26.23	14.98	17.55	8.93	7.75
1-6	52	67	4.29	11.94	7.31	21.51	10.35	14.97	6.18	7.05

Experience Groups: I = Low Experience Scores; II = High Experience Scores.

tained from only 119 of the 212 children taking the test. The data again shows consistently higher means for Group II and it is particularily interesting to note that for RL scores, the σ's for this group are considerably larger than those for Group I. This does not hold true for σ's of the "mean correct" scores. The "t" tests for the significance of the differences between the means of these two groupswould appear to indicate that such experience is not reflected in a task which involves the learning of musical material from auditory stimulus alone.

Grade	"t" (RL)	"t" (Mean Correct)	Grade	"t" (RL)	"t" (Mean Correct)
1, I II	1.099	1.240	5, I II	1.32	1.833
2, I II	2.105	2.994***	6, I II	0.558	0.547
3, I II	1.174	1.687	1-6, I II	2.438**	3.702***
r, I II	1.184	0.623			

* Significant at the 5% level.
** Significant at the 2% level.
*** Significant at the 1% level.

These tests show that for both kinds of scores the hypothesis of no difference between experience groups is untenable only when all the grades are combined, the single exception occurring at the second grade level. This is considerably different from the findings of similar "t" tests for the 45-Item Test where the hypothesis could be rejected at every grade level considered. It is evident from this that other hypotheses regarding the relationship of experience and musical learning will need to be tested.

Summary of the Findings

The results of the Phrase Test which had been given to 212 children showed the following:

1. The "t" tests do not completely justify rejecting the hypothesis of "no difference" between the means of two grades in 11 of the 12 comparisons made for the RL scores. The relationship of grade level to the learning of a musical phrase, such learning based upon auditory perception alone, is apparently different from the relationship between grade level and the auditory perception of single tonal configurations.
2. The "t" tests made for "mean correct" scores justify rejecting the hypothesis of "no difference" between the means of two grades in 5 of the 12 comparisons made. These scores, since they represent achievement and do not directly reflect the degree of learning taking place, give results similar to those for the 45-Item Test.
3. The correlations between the two types of scores were, except for grade 1, significant at the .01% level indicating a high relationship between the rate of learning and achievement.
4. The correlations between the RL scores and the 45-Item Test scores are, except for the first and third grade, significant at or beyond the 1% level; those for the "mean correct" scores are all significant beyond the .01% level. These clearly show that there is a significant relationship between the two tasks.
5. The "t" tests continued to show that the hypothesis of "no difference" between the means of boys and girls is tenable.

6. The "t" tests do not warrant rejecting the hypothesis of "no difference" between the means of the two experience groups, except for grade 2 and for the six grades combined. This suggests that musical experience and environment is not significantly related to the task of learning a musical phrase, that other factors are present which lead to success or failure.

20-Item Test

The primary function of this test was to ascertain the degree to which children gave consistent responses to given auditory stimuli. The twenty items, selected on the basis of the item analysis of the Exploratory Test, were randomized for each of three successive trials. It was not the intent of this test to determine the number of trials needed to obtain a consistently accurate response and so successive presentations of the same item were avoided. Usable data was obtained for 164 children, 36 fewer than the number who had been scheduled for the test. The items were scored exactly as they had been scored for the 45-Item Test, with the total test score representing the sum of the scores for each trial divided by three. Table XXIII summarizes these scores.

The data shows that there is relatively little difference between the means for the several trials at each grade level except for the 4th and 6th grades. The hypothesis of "no difference" between the means of the several grades being compared could not be rejected except for grades 5 and 6. These "t" tests were:

TABLE XXIII

Summary, by Grade, of Scores on the 20-Item Test

Grade	N	Mean				σ			
		Total Test*	Trial I	Trial II	Trial III	Total Test*	Trial I	Trial II	Trial III
1	34	104.7	102.7	105.6	106.5	52.9	52.3	54.3	54.3
2	29	114.3	113.1	116.2	113.4	47.7	47.5	47.4	50.5
3	28	113.5	114.4	113.7	112.3	53.0	50.7	54.9	55.2
4	31	130.7	129.3	131.3	131.8	53.7	54.2	54.8	53.5
5	24	109.8	104.7	111.8	112.9	48.0	46.8	47.4	51.6
6	18	143.6	143.1	142.7	145.2	47.1	46.0	48.5	48.1
1-6	164	117.8	116.3	118.7	118.7	52.4	51.8	53.1	54.1

* Mean score of the three trials

Grades	"t" (Total Test)	"t" (Trial I)	"t" (Trial II)	"t" (Trial III)
1 and 2	0.734	0.810	0.800	0.509
2 and 3	0.060	0.092	0.181	0.074
3 and 4	1.221	1.069	1.211	1.355
4 and 5	1.472	1.738	1.363	1.300
5 and 6	2.221*	2.591**	2.018	2.020*
1 and 3	0.635	0.871	0.567	0.410
2 and 4	1.231	1.201	1.121	1.349
3 and 5	0.252	0.698	0.129	0.036
4 and 6	0.828	0.892	0.715	0.858
1 and 4	1.934 almost 5%	1.978	1.865	1.864
2 and 5	0.330	0.638	0.329	0.035
3 and 6	1.924 almost 5%	1.904	1.787	2.027*

* Significant at the 5% level.
** Significant at the 2% level.

The results continue to agree, generally, with the findings for the other tests. The "t" values for grades 1 and 4 and grades 3 and 6 very closely approach the 5% level of significance and so the hypothesis of "no difference", when considering a three-year interval, may not be retained with complete confidence.

The mean scores for the total test show that when boys and girls are compared, these comparisons agree with the findings for the other tests.

The data is summarized in Table XXIV, with the "t" tests given as follows:

Grade	"t"	Grade	"t"
1, BG	0.281	5, BG	0.198
2, BG	2.537*	6, BG	0.529
3, BG	0.207	1-6, BG	1.316
4, BG	0.865		

* Significant at the 5% level.

This is the first test for which an obtained "t" has met the criterion of significance and, in view of the evidence which has thus far been presented, the writer feels that this is quite probably the result of chance and would not reject the hypothesis.

Data regarding their musical experience was secured from only 79 of the 164 subjects taking the 20-Item Test. The small number in each group at each grade level creates certain problems of interpretation. This is summarized in Table XXV and the limited number of cases in grade 6 tends to make the "t" test meaningless. The "t" values obtained were:

Grade	"t"	Grade	"t"
1	5.242**	5	1.011
2	1.273	6	20.000**
3	4.760**	1-6	5.190**
4	2.262*		

* Significant at the 2% level.
** Significant at the .01% level.

The hypothesis may be rejected for all but grades 2 and 5 and, as for the the 45-Item Test, those children with above-average experience performed significantly better than those with limited experience. There are, of course, reservations which need to be made in view of the small number of

TABLE XXIV

Summary by Sex, of Scores on the 20-Item Test

Grade	N		$\bar{X}$		σ	
	B	G	B	G	B	G
1	21	13	102.7	108.1	56.6	46.3
2	7	22	77.1	126.1	25.1	47.2
3	13	15	111.2	115.5	46.1	58.2
4	16	15	122.5	139.5	57.1	48.2
5	11	13	107.6	111.7	49.3	46.7
6	9	9	149.8	137.4	35.8	55.5
1-6	77	87	112.1	122.9	52.8	51.5

TABLE XXV

Summary, by Experience Groups, of Scores on the 20-Item Test

Grade	N		$\bar{X}$		σ	
	I	II	I	II	I	II
1	8	5	84.0	164.0	30.6	9.1
2	8	7	87.3	119.4	52.9	35.1
3	9	8	61.3	155.6	33.5	43.1
4	10	4	100.6	168.0	50.0	36.9
5	7	9	92.6	119.0	44.6	51.4
6	2	2	160.0	200.0	2.0	0
1-6	44	35	88.5	144.1	47.1	46.1

Experience Groups: I = Low experience score;
II = High experience score.

children involved in these comparisons.

The 20-Item Test was, of course, very closely related to the 45-Item Test since the content was identical although not as extensive. The correlations between scores for the two tests, using the mean score for all three trials of the 20-Item Test, were all significant far beyond the 1% level:

Grade	"r"	Grade	"r"
1	.94	5	.92
2	.95	6	.94
3	.82	1-6	.92
4	.96		

These results are to be expected when the performance on two almost identical tasks are compared.

The primary function of this test was to ascertain the consistency with which children responded to the same stimulus. The items had been randomized for each of the three trials to insure that the response resulted from attention to the stimulus. Had the subject known that he was to be given successive trials of the same item his responses undoubtedly would have reflected such awareness. The data summarized in Table XXIII, page 93, shows that the means and σ's remain relatively unchanged from trial to trial. This suggests that at all grade levels children consistently gave the same kind of response to each presentation of the item. If their responses were markedly different each time the item was presented one would have to conclude either that they were not concentrating on the task or that a given stimulus sounded different from one trial to the next. The similarity of means and σ's also suggests that the practice of the first trial had little effect upon the score obtained for the second trial and that the practice from the first two trials likewise had little effect upon the score of the third trial.

The "t" tests made between each pair of trials, I-II, I-III, and II-III, showed the differences between the means to be significant between the 85% and 90% levels, with many of them approaching the 99% level. The hypothesis that consistent responses, as reflected by consistent scores from trial to trial, obviously may not be rejected for this sample. The "t" tests were as follows:

	"t"	"t"	"t"
Grade	I-II	I-III	I -III
1	0.224	0.289	0.064
2	0.239	0.0.8	0.213
3	0.047	0.141	0.091
4	0.144	0.186	0.039
5	0.513	0.565	0.074
6	0.027	0.131	0.154
1-6	0.415	0.410	0.001

The raw data for this test shows several interesting things which are not readily apparent from the means and σ's. A detailed analysis was made of the performance on each item for each trial of the test. This analysis was concerned with identifying the following:

a. The number of items which received the same response for all three of the trials, or for two of the three trials. A separate tabulation was made for items which were consistently correct and those which were consistently incorrect.

b. The number of items which received correct responses for all three trials; for two of the trials; for only a single trial; or were not correct on any trial.

The results were tabulated in terms of the kinds of responses made for a given item. The responses were: (1) not correct for any of the three trials; (2) correct for either one, two, or three of the three trials; and (3) the same, but incorrect, for two or three of the three trials. The total item count for any grade level was obtained by multiplying the number of test items, 20, by the number of subjects. The data in Table XXVI summarizes these several kinds of responses as percentages of the total item count.

The data under Part A of the Table shows, except at the fifth grade level, that the number of items for which correct responses are made in three of the three trials increases from 25% of the total item count at first grade to 50% of the total item count in sixth grade. This is accompanied by a decrease in the number of items for which incorrect responses are made in all three trials; 54% of the total item count at the first grade level and only 32% of the total item count at the sixth grade level. The number of items receiving correct responses in one or two of the three trials does not change very much from grade level to grade level. This shows the 4th and 6th grade children of the sample to perform much more consistently, in terms of correct responses, than children in the other grades. However, many children will perform consistently but this performance may well be inaccurate.

The data in Part B of the table shows that when we examine items for which the same response is made in two of the three trials, there is an

TABLE XXVI

Summary of Responses to the 20-Item Test

A. Percentage Distribution, by Trials, of Correct Responses

Grade	N	Correct in 1 of 3 trials	Correct in 2 of 3 trials	Correct in 3 of 3 trials	Item is incorrect in all 3 trials
1	34	11%*	10%	25	54%
2	29	11	10	29	50
3	28	12	9	32	47
4	31	9	8	43	40
5	24	12	12	27	49
6	18	7	11	50	32
1-6	164	11	10	33	46

* 11% of the total item count for this grade were items which were correct only on one of the three trials.

B. Percentage Distribution, by Trials, of Consistent Responses

Grade	N	2 of 3 Trials		3 of 3 Trials		Total of Cols. 3 to 6
		Correct	Incorrect	Correct	Incorrect	
1	34	10%**	16%**	25%	3%	54%
2	29	10	14	29	5	58
3	28	9	9	32	3	53
4	31	8	11	43	2	64
5	24	12	12	27	1	52
6	18	11	11	50	1	73
1-6	164	10	12	33	3	58

** In two of the three trials, 10% of the total item count were correct and 16% were the same but incorrect.

almost equal division between correct and incorrect responses at each grade level. This does not hold true for items having identical responses in three of the three trials, since very few items are consistently incorrect. The final column in Part B of the table gives a summary of the consistency with which children make the same response to an auditory stimulus. Over half of the total item count for each grade showed that the item received the same response in two or three of the three trials. This percentage is much higher for grades 4 and 6. The total group performs at about a 60% level of consistency and the hypothesis that several presentations of a given auditory stimulus will usually result in identical responses appears to be tenable.

Summary of the Findings

1. The "t" tests did not warrant rejecting thehypothesis of "no difference" between the means of two grades except for grades 5 and 6. Comparisons between grades taken at three year intervals did not result in differences which were significant.
2. The data shows that the boys and girls performed at approximately the same level for this task.
3. The differences between the two experience groups were, excluding the 6th grade because of incomplete data, highly significant in four of the six comparisons.
4. The correlations between this test and the 45-Item Test show that the two tasks result in almost parallel performance.
5. The "t" tests between trials showed that correct responses to the items remained consistent from one trial to the next.

6. This consistency of response is also evident in the analysis of the raw data. This showed that from 52 to 73 per cent of the items in the total item count had identical responses in two or three of the three trials.

CHAPTER V

SUMMARY AND CONCLUSIONS

Procedures

This study has been primarily concerned with the auditory perception of musical sounds by children in the first six grades. A stratified sample of over 600 children randomly selected at each of the first six grade levels was used in the study. These children were drawn from the total elementary school population of the public elementary schools in Madison, Wisconsin.

Four tests were constructed for this study, the content of these tests based upon an extensive analysis of tonal and rhythmic configurations found in the songs children sing. The 45-Item, 20-Item and Rhythm Tests were designed to collect data regarding the auditory perception of common tonal configurations. For each of these tests, the subject heard an aural presentation of the tonal configuration and then was asked to sing an immediate response to the stimulus. The accuracy of this response was then interpreted as a measure, not only of vocal control, but of the level at which auditory perception functioned. The scoring procedures recognized, and rewarded, several types of responses:

a. The response, because it was a complete or partial duplication of the stimulus, demonstrated that auditory perception was functioning at a high level and also that the subject possessed excellent vocal control.

b. The response, while it accurately reflected the general shape and content of the stimulus, was not a duplication of the stimulus and thus showed that auditory perception was not functioning at as high a level as observed in the Type 1 response. The accuracy of singing showed, however, that poor vocal control was not necessarily a contributing factor.

c. There was evidence of a partial awareness of the nature of the stimulus in that the response contained _either_ the same number of tones as the stimulus _or_ followed the general direction of the stimulus. The absence of both characteristics of the stimulus indicated that auditory perception was functioning at a relatively low level and that poor vocal control might also be present in the situation.

The Phrase Test was designed to collect data regarding the auditory perception of a more complete musical entity. Success in learning a short phrase from aural presentations alone depended not only upon vocal control, but demanded that the subject be able to retain the aural impression of the stimulus so his response could be compared against the stimulus.

All of the subjects were given the 45-Item Test. The total sample was then randomly assigned to one of three groups and each group then given one of the other three tests. Certain information was obtained from most children regarding their own musical training and the nature and extent of musical activity in the home. Each child was tested individually throughout the investigation, with these responses tape-recorded for subsequent processing and scoring. While this individualized testing proved to be very time-consuming, it does permit the collection of certain kinds of data which could not be obtained any other way. Each of the tests was tape-recorded to permit uniformity of testing procedures.

Summary of Major Findings Based on the Data

1. There are no significant differences between boys and girls in terms of the auditory perception of musical sounds. Only for one of the 35 comparisons did the mean score of the boys differ significantly from the mean score of the girls.
2. The hypothesis that age, as defined in terms of grade level, is a significant factor in the development of auditory perception, may be accepted with certain reservations. The data showed that when the two grades being compared are taken at one-year intervals, such as grades 1 and 2, the differences between the means are significant in only six of 25 comparisons. Four of these six were for grades 5 and 6. When a two-year interval is used, such as grades 1 and 3, the differences between the means are significant in five of 20 comparisons, with three of these for grades 4 and 6. Finally, when a three-year interval is used, this interval clearly comparing the lower and upper grades, the differences between the means are significant for seven of the 15 comparisons, with four additional comparisons very closely approaching the 5% level of significance.
3. Competence levels established for the 45-Item Test indicate that the proportion of children earning relatively high scores tends to increase at each successive grade level. On the other hand, the proportion of children earning scores within each of the other three competence levels tends to remain relatively constant from grade to grade, except for the

sixth grade. These findings tend to support the hypothesis that grade level and the development of auditory perception are related if one limits such comparisons to the above-average competence groups. The data does not, and cannot, indicate this relationship for the other competence levels since it does not reflect changes which might take place in the individual child as he moves from grade level to grade level.

4. The item analysis, made in terms of the type of responses given to the 45-Item Test, showed that more than half of the children at each grade level were able to give correct or partially correct responses. Approximately 25 per cent of the children at each grade level showed an awareness of the shape and/or content of the item. These responses, because they were in tune, indicated a high level of vocal control and despite this the children were unable to duplicate the stimulus. Finally, the proportion of children whose responses indicated that they were seemingly unaware of the nature of the stimulus and were lacking in good vocal control also tended to remain constant at each grade level, except for the sixth. This, together with the information regarding competence levels and the observed differences between the test means for the several grades, makes it difficult to accept or reject hypotheses one might make regarding the relationships between age and the development of auditory perception. However, good vocal control undoubtedly reflects good auditory perception and vice versa.

5. The correlations between the 45-Item Test and the Rhythm and 20-Item Tests are all uniformly high, indicating consistency of performance from one task to the next. ThePhrase Test, because both the RL and "mean correct"

scores reflect the more complex nature of the task, yield lower correlations with the 45-Item Test, although they are all significant. This suggests that the auditory perception of short tonal configurations does, for many individuals, precede the auditory perception of larger musical units. It is expected that data secured for the same children over a period of several years might provide a better basis for identifying this kind of relationship.

6. The analysis of data for the Phrase Test showed that the scores which identified the rate at which children were able to learn the phrase did not reflect the variables of age and musical experience nearly as much as did scores which identified overall achievement. While this is true when the group means are compared, the data also shows that more children in the upper grades demonstrated the ability to learn the phrase than was observed for children in the lower grades.
7. The hypothesis that musical training and experience is a significant factor in the development of auditory perception may not be rejected with confidence since significant differences between the means of these two groups were obtained in 17 of 35 comparisons. Three additional comparisons gave results which closely approximated the 5% level of significance and nine of the remaining 15 comparisons were at, or beyond the 25% level. The 45-Item Test, because data regarding the variable was available for a substantial number of children, gave results which clearly supported this hypothesis; the "t's" were significant at each grade level.
8. The hypothesis that the addition of the rhythmic element to a melodic pattern has a significant influence upon the auditory perception of such

melodic patterns may be rejected since none of the "t" values approached the 5% level of significance. This indicates that the presence of rhythm apparently does not influence the auditory perception of melodic items.

9. The data for the 20-Item Test showed that at all grade levels, children are capable of making consistent responses to a given auditory stimulus. From 52 to 73 per cent of the items received the same response in two or three of the three trials. The mean scores for each of the several trials at each grade level also reflected this consistency since none of the "t" values met the criterion of the 5% level of significance.

10. Throughout the study it is evident that the 6th grade sample may well have been the result of randomly selecting a large number of musically superior children. It is also possible in view of poorer performance of the 5th grade group, that this might have been the result of randomly selecting a large number of children with below-average musical ability. For both groups the probability that this sample contains a normally distributed population in terms of musical ability is somewhat suspect. However, without additional data there is no way of ascertaining the true nature of either group. It is quite probable that had all [illegible] children been given all four of the tests, and had data regarding musical experience been obtained for the total sample, the findings of this study might have been somewhat different.

Implications of the Study

1. There is considerable evidence that age and auditory perception are

closely related if one compares the lower grades with the upper grades. This relationship is not as evident when grades are compared at two-year intervals, and even less evident if a one-year interval is taken. This suggests that the observed differences between lower and upper grades may result more from the music program in the school than from age alone. This would need further study and in order to suggest possible approaches and new hypotheses, the data from this investigation will be examined using other kinds of statistical analyses.

2. The instrument used to gather data necessary for identifying the kinds and amounts of musical experience needs to be refined and expanded. It is also evident that complete data must be secured for every child in the sample. Despite the obvious weaknesses, the data shows the musically experienced child to have a decided advantage in responding to the auditory presentations of musical material. The child not only is more able to control the singing voice but has a better understanding of musical material and, because of this understanding, is able to successfully compare his response to the stimulus and make judgments regarding the accuracy of such a response.
3. The existence within each grade level, of children with marked differences in terms of musical competence, auditory perception, and aural understanding further emphasizes the need for developing teaching procedures and activities which will take account of these differences and make for more effective learning.
4. The data from the present study shows that the proportion of children earning scores in one of the three lower competence levels tends to re-

main constant from one grade level to the next. The data also shows that the proportion of children making responses which indicate either a partial understanding or a complete lack of aural understanding also tends to remain constant from grade level to grade level. The writer plans to study the possible reasons for this apparent lack of change since he is reluctant to accept the hypothesis that the music program of the schools fails to contribute to the musical growth and development of the child. To accomplish this, the first, second and third grade children who participated in this study will be re-tested each year until each group is sixth grade. This longitudinal study is expected to give data which will permit the testing of several related hypotheses concerning the musical development of children.

5. Although the 45-Item Test was highly reliable it is evident that other kinds of tasks need to be developed which will measure the auditory perception of musical sounds without reference to responses which require singing ability. The data from the present study suggests that the development of auditory perception parallels the development of vocal control, since the kinds of responses which showed poor vocal control also failed to reflect the nature of the stimulus. It is entirely possible, however, that lack of vocal control, as shown by the "uncertain singer" does not necessarily reflect a corresponding lack of aural understanding. To fully test such hypotheses would require a careful study of this particular kind of singer, using specially designed tests and testing procedures.
6. The results of the Rhythm Test showed that a combination of two musical

elements apparently had little influence upon the auditory perception of a single musical element. The writer plans to study this aspect of auditory perception in some detail since it is assumed that complex stimuli are more difficult to perceive than simple stimuli.

7. The findings of the study suggest that the music program of the schools must, if the children are to develop an aural understanding of musical sounds, include a variety of activities which are designed to stimulate and challenge the child. Rote learning of songs frequently fails to give the child an opportunity to become independent; to develop the ability to judge the accuracy of his response as compared to the stimulus; or to become aware of the subtle differences between similar but not identical tonal patterns. Aural understanding, which is a reflection of accurate auditory perception, results from thinking and not mechanical imitation; from judgments made independently by the child and based upon his understanding of basic musical concepts and not judgments made for the child by someone else. Music educators need to examine the content of the music program, not only with reference to the kinds of attitudes this program seeks to develop, but also in terms of the kinds of musical skills and competencies which give meaning and purpose to such a program.
8. The research techniques and procedures employed in the study are reliable and adequate, but very inefficient. If instruments can be developed which will permit effective and reliable testing of auditory perception in group situations, such instruments would undoubtedly be widely used in the schools. There is a widespread need for efficient and objective measures of the several aspects of musical competence and understanding.

BIBLIOGRAPHY

1. Bugg, Eugene G. and Herpel, Lloyd "The Significance of Tonal Memory for Musicality", *Journal of General Psychology*, 35:3-15 (July '46).

2. Drexler, Edith N. "A Study of the Ability to Carry a Melody at the Pre-School Level", *Child Development*, 9:319-332 (Sept. '38).

3. Hatwick, Melvin S. "The Role of Pitch Level and Pitch Range in the Singing of Pre-School, First Grade, and Second Grade Children", *Child Development*, 4:281-291 (1933).

4. Heinlein, Christian Paul "Brief Discussion of the Nature and Function of Melodic Configuration in Tonal Memory, with Critical Reference to The Seashore Tonal Memory Test", *Journal of Genetic Psychology*, 35:45-61 (1928).

5. Jersild, Arthur T. and Bienstock, Sylvia F. "The Influence of Training on the Vocal Ability of 3-Year Old Children", *Child Development*, 2.272-291 (1931).

6. Jersild, Arthur T. and Bienstock, Sylvia F. "A Study of the Development of Children's Ability to Sing", *Journal of Educational* Psychology, 25:481-503 (Oct. '34).

7. Jersild, Arthur T. and Associates *Training and Growth in the Development Children*, Child Development Monographs, No. 10, 1932. (Chapter 8)

8. Ortmann, Otto *On The Melodic Relativity of Tones*, Psychological Monographs, Vol. 35, No. 1 (Whole No. 162), 1926.

9. Ortmann, Otto "Some Tonal Determinants of Melodic Memory", *Journal of Educational Psychology*, 24:454-467 (Sept. '33).

10. Petzold, Robert G. "The Perception of Music Symbols in Music Reading by Normal Children and by Children Gifted Musically", *Journal of Experimental Education*, 28.271-319 (June '60).

11. Ross, C. C. and Stanley, Julian C. *Measurement in Today's Schools* New York. Prentice-Hall Inc., 1954.

12. Updegraff, Ruth; Heiliger, Louise; and Learned, Janet "The Effect of Training upon the Singing Ability and Musical Interest of 3-, 4-, and 5-year Old Children", *University of Iowa Studies in Child Welfare*, 14:83-131 (1938).

13. Williams, Harold; Sievers, C. H.; and Hattwick, Melvin *The Measurement of Musical Development*, University of Iowa Studies in Child Welfare, Vol. 7, No. 1, 1933.

APPENDIX A

Song-Texts Analyzed for Rhythmic Patterns

1. *The American Singer*, Second Edition
   New York: American Book Company, 1955

   Books 4 and 6

2. *Music For Living*
   New York: Silver Burdett Co., 1956

3. *Music for Young Americans*
   New York: American Book Co., 1959

4. *New Music Horizons*
   New York: Silver Burdett Co., 1946

   Books 2, 4 and 6

5. *Our Singing World*
   New Cork. Ginn and Co., 1951

   Book 2, *Singing on Our Way*
   Book 4, *Singing Every Day*
   Book 6, *Singing in Harmony*

6. *Together We Sing*
   Chicago. Follett Pub. Co., 1958

   Book 2, *Music Round the Town*
   Book 4, *Music Across Our Country*
   Book 6, *Voices of the World*

APPENDIX B

INFORMATION SHEET

Please answer all of the questions on this sheet as carefully as you can. Thank you for your help.

- -

1. My name is ____.

2. I am in ____________grade this year. I was born________________.
Month Day Year

3. I play the ________________ [Give the name of the musical instrument] and took lessons for ________ years.

4. I also play the ________________[Give the name of the other instrument] and took lessons for ________ years.

5. My father plays the ____________. [Do not answer if your father does not play a musical instrument now.]

6. My mother plays the____________. [Do not answer if your mother does not play a musical instrument now.]

7. I hear my parents, or just my mother or father, play their instruments at home: [Put a check mark in front of one of these]

_____Never _____ Once in a while ______ Often

8. My father sings in a church choir, or some other chorus. —— True——False

9. My mother sings in a church choir or some other chorus. ——True——False

10. I hear my mother or my father sing songs at home. [Check one of these]

_____ Never _____ Once in a while ________ Often

11. I have older brothers or sisters who play the ____________________.
[Do not answer if your older brothers or sisters do not play some musical instrument.]

12. I have younger brothers or sisters who play the__________________.
[Do not answer if your younger brothers or sisters do not play some musical instrument.]

13. My parents have a record collection. ___True _____False

14. In our home we listen to records: [Check one of these]

______ Never ____ Once in a while _____ Often

APPENDIX B

Information Sheet

(Scoring Key

1.)
) Do not score.
2.)

3. 5 points if an instrument is named; *plus* 3 points for each year studied.

4. 5 points if an instrument is named; *plus* 3 points or each year studied.

5. 5 points if an instrument is named.

6. 5 points if an instrument is named.

7. Never = 0; Once in a while = 5; Often = 10.

8. True = 5; False = 0.

9. True = 5; False = 0.

10. Never = 0; Once in a while = 5; Often = 10.

11. 5 points if an instrument is named.

12. 5 points if an instrument is named.

13. True = 5; False = 0.

14. Never = 0; Once in a while = 5; Often = 10.

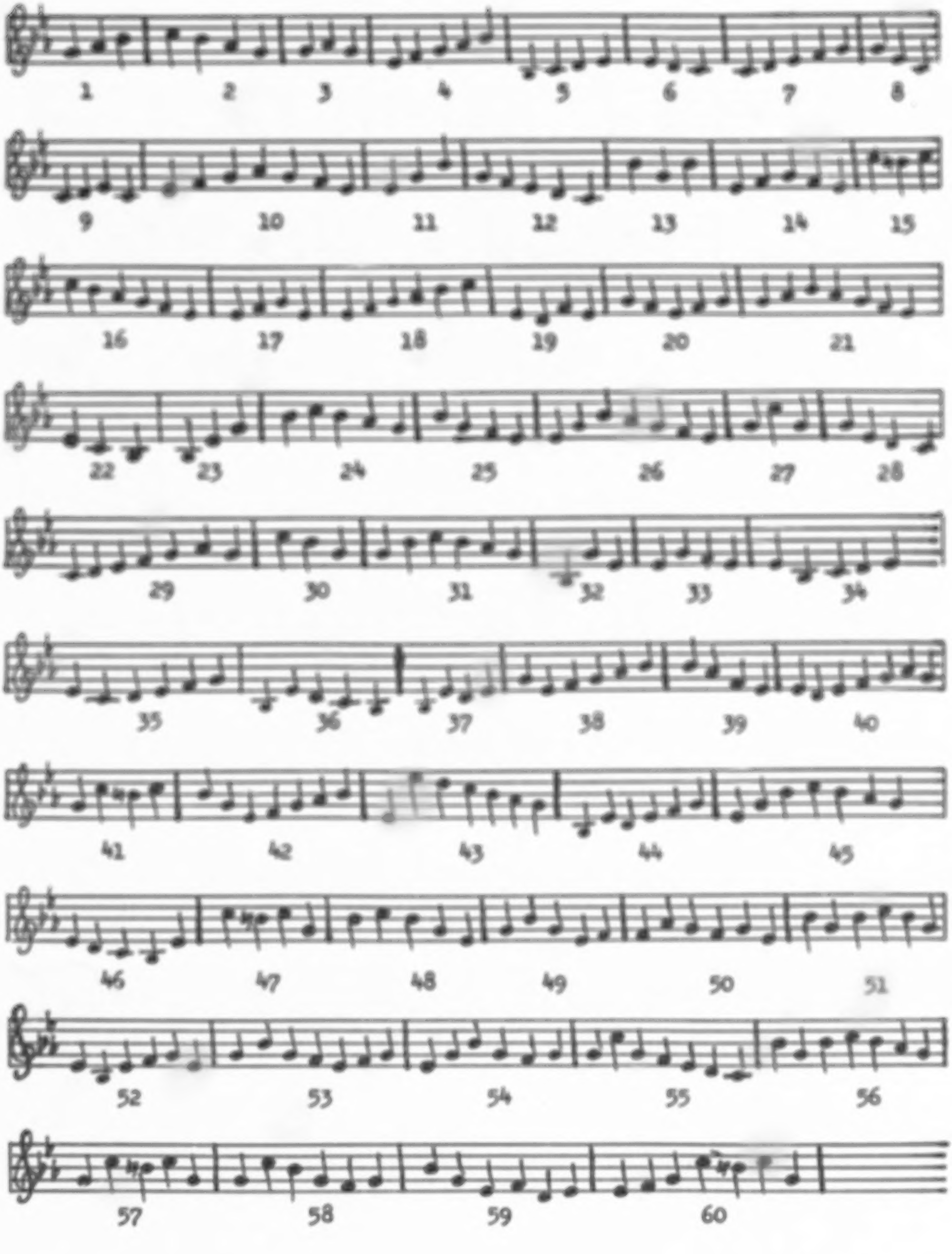
APPENDIX B

Summary of Responses to Information Sheet by 506 children

	No		Yes	
Item	N	%	N	%
	399	78.8	107	21.2
4	496	98.0	10	2.0
5	440	87.1	66	12.9
6	409	80.8	97	19.2
	485	95.8	21	4.2
9	467	92.3	39	7.7
11	373	73.6	133	26.4
12	488	96.4	18	3.6
13	150	30.0	356	70.0

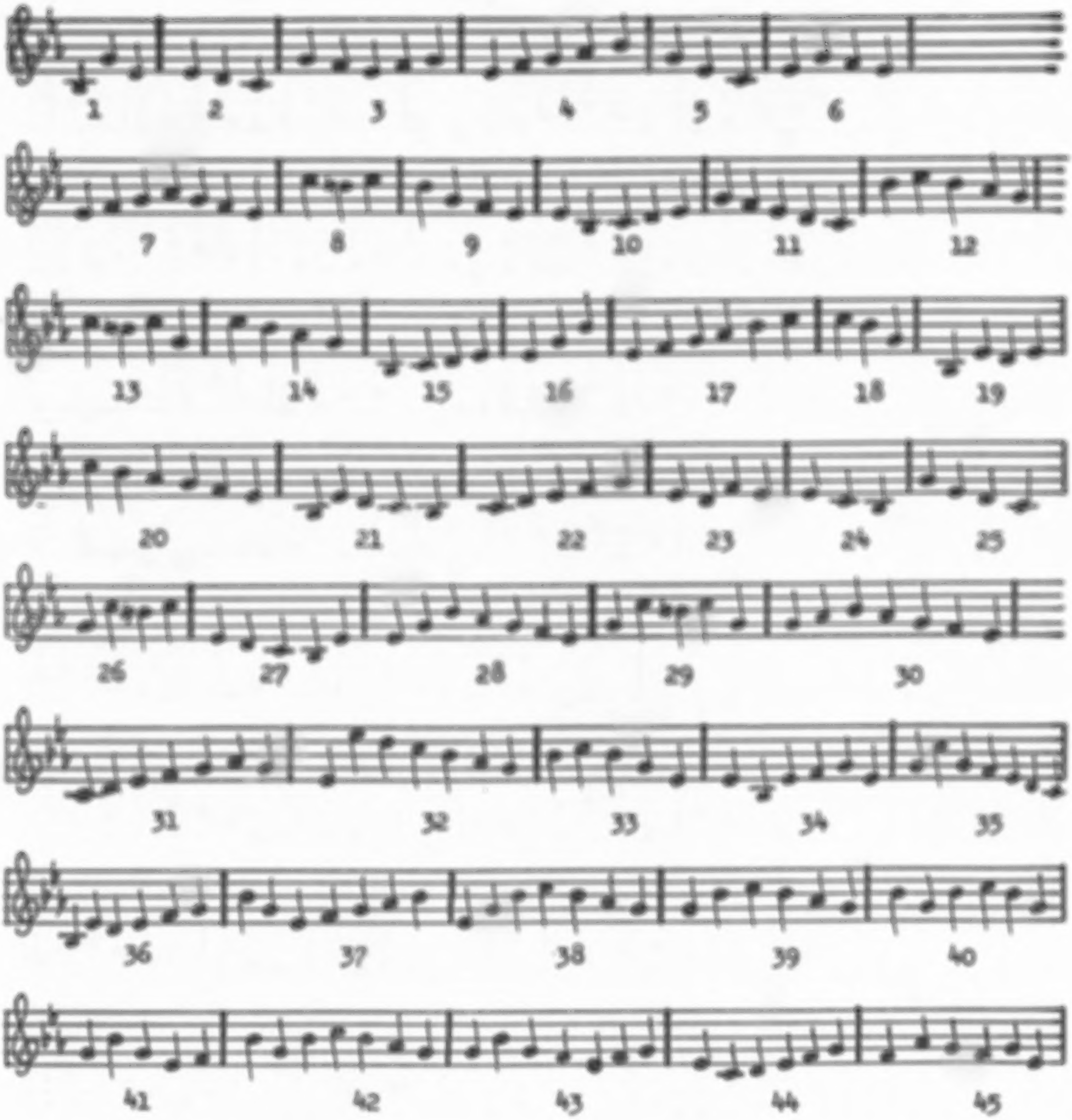
	Never		Once in a While		Often	
Item	N	%	N	%	N	%
7	377	74.6	93	18.3	36	7.1
10	48	9.5	287	56.6	171	33.7
14	44	8.7	215	42.5	247	49.8

APPENDIX C, 60-Item Test


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

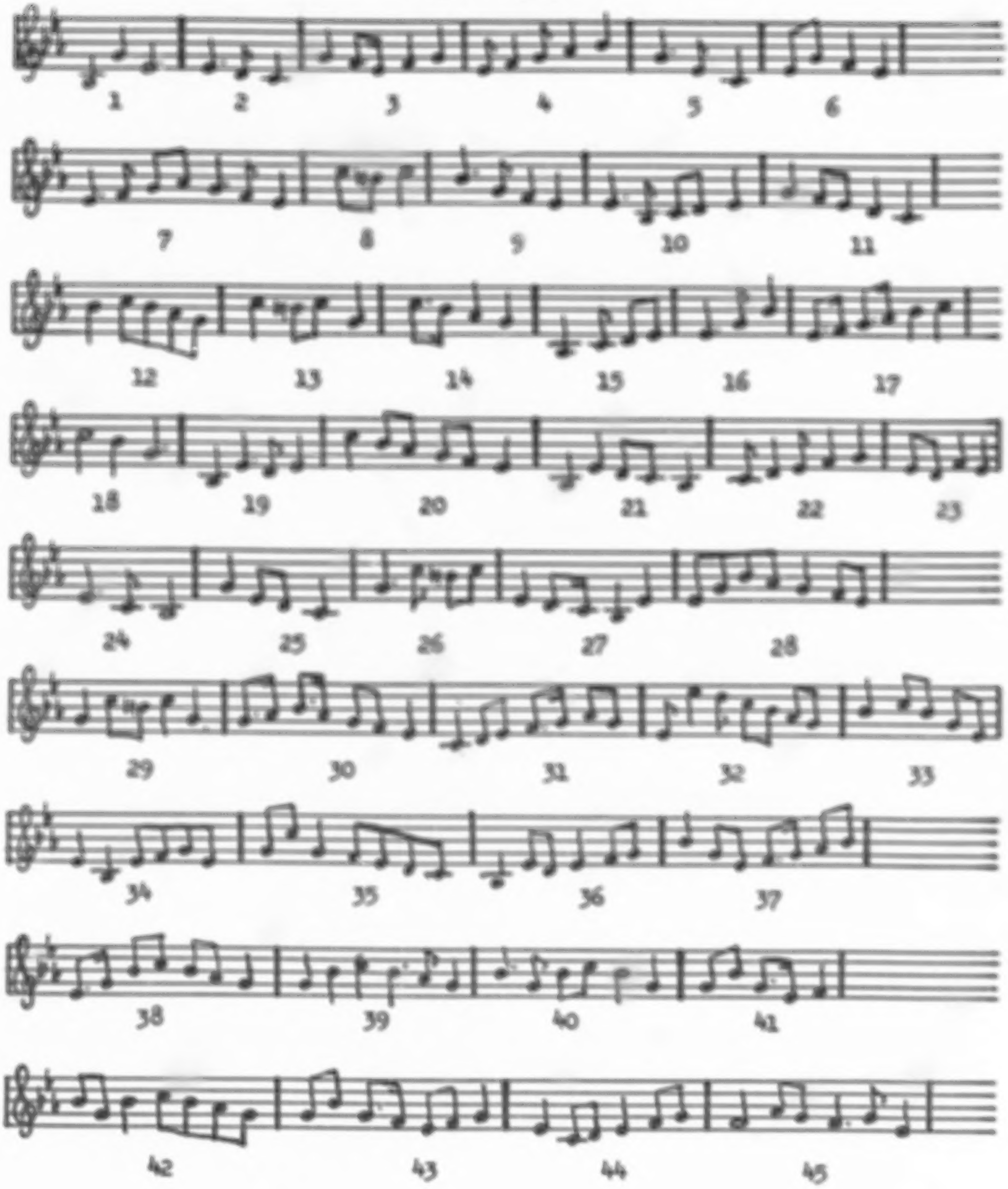
APPENDIX C

45-Item Test


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

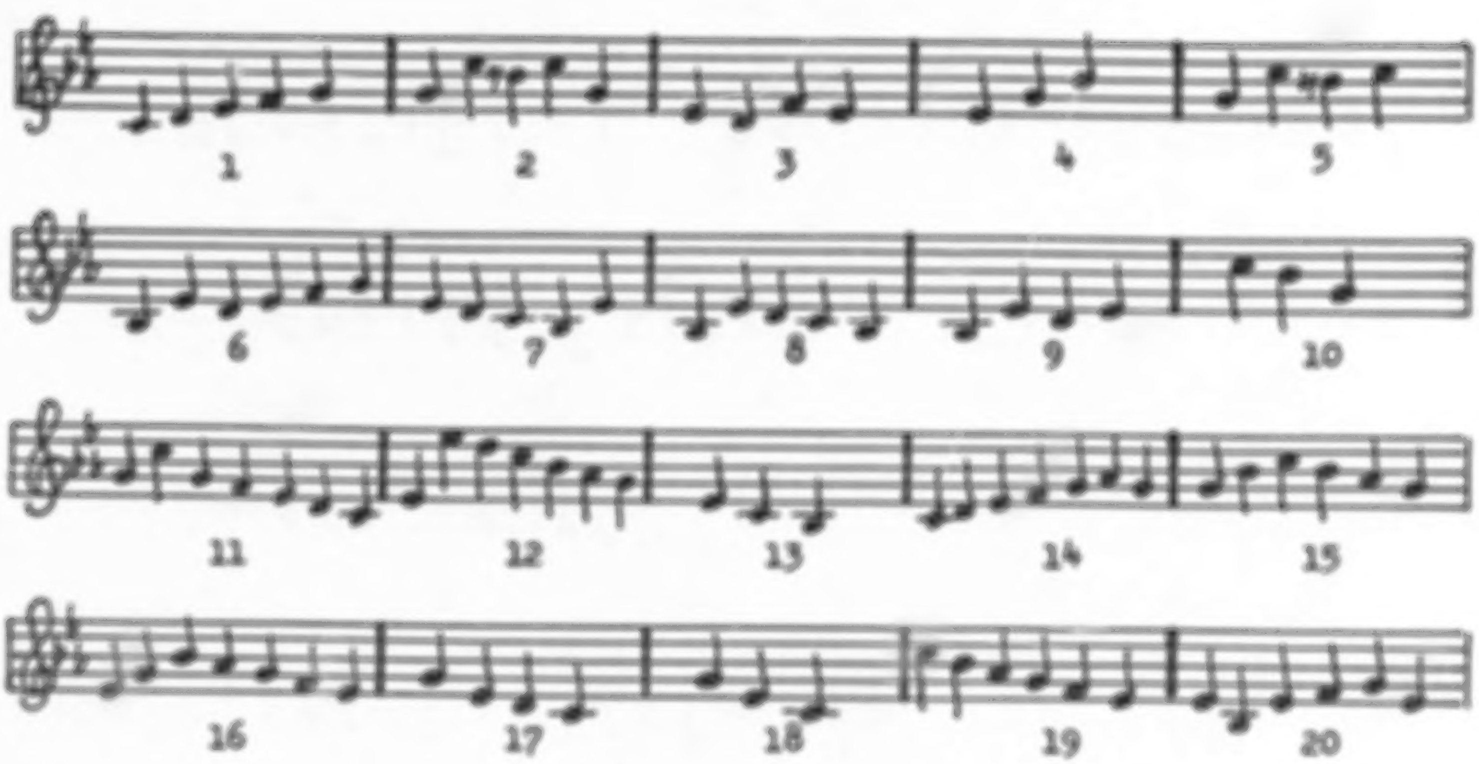
APPENDIX C

Rhythm Test


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

APPENDIX C

20-Item Test, One Trial Shown


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Phrase Test